Simultaneous enforcement of constraints at position and velocity levels in the nonsmooth generalized-α scheme

Olivier Brüls[a,*], Vincent Acary[b], Alberto Cardona[c]

[a] *University of Liège, Department of Aerospace and Mechanical Engineering (LTAS), Chemin des Chevreuils, 1 (B52), 4000 Liège, Belgium*
[b] *INRIA Rhône-Alpes, Centre de recherche Grenoble, 655 avenue de l'Europe, Inovallée de Montbonnot, 38334 St Ismier Cedex, France*
[c] *Universidad Nacional del Litoral - Conicet, CIMEC, Colectora Ruta Nac 168 / Paraje El Pozo, 3000 Santa Fe, Argentina*

Abstract

This paper presents a formalism for the transient simulation of nonsmooth dynamic mechanical systems composed of rigid and flexible bodies, kinematic joints and frictionless contact conditions. The proposed algorithm guarantees the exact satisfaction of the bilateral and unilateral constraints both at position and velocity levels. Thus, it significantly differs from penalty techniques since no penetration is allowed. The numerical scheme is obtained in two main steps. Firstly, a splitting method is used to isolate the contributions of impacts, which shall be integrated with only first-order accuracy, from smooth contributions which can be integrated using a higher order scheme. Secondly, following the idea of Gear, Gupta and Leimkuhler, the equations of motion are reformulated so that the bilateral and unilateral constraints appear both at position and velocity levels. After time discretization, the equations of motion involve two complementarity conditions and it can be solved at each time step using a monolithic semi-smooth Newton method. The numerical behaviour of the proposed method is studied and compared to other approaches for a number of numerical examples. It is shown that the formulation offers a unified and valid approach for the description of contact conditions between rigid bodies as well as between flexible bodies.

Keywords: nonsmooth contact dynamics, flexible multibody system, time integration, generalized-α method, time-stepping schemes, index reduction

1. Introduction

This paper studies numerical algorithms for the simulation of mechanical systems including rigid and flexible bodies, kinematic joints and frictionless contact

*Corresponding author, phone: +32.4.366.91.84, fax: +32.366.92.17
Email addresses: o.bruls@ulg.ac.be (Olivier Brüls), vincent.acary@inria.fr (Vincent Acary), acardona@intec.unl.edu.ar (Alberto Cardona)

conditions. The condition of impenetrability of the bodies in contact is expressed as a unilateral constraint, with the consequence that impacts and/or instantaneous changes in the velocities may arise in the dynamic response. The design of a consistent and stable time integration scheme for such systems requires great care.

If dynamic contacts are analyzed between elastic solids and structures, the gap velocity is necessarily discontinuous otherwise a non-physical penetration of the bodies would occur. Due to the flexibility of the bodies, the contact pressure remain nevertheless finite and no impulsive force is observed in this case. Standard schemes for mechanical systems, such as the Newmark, HHT or generalized-α methods, are based on the assumption that displacements and velocities evolve smoothly as a function of time and are thus unable to deal consistently with velocity jumps. This motivated the development of energy-consistent schemes or stabilization techniques [1, 2, 3] for the analysis of dynamic contacts between flexible bodies.

When contacts between rigid bodies are studied, e.g., in multibody systems, impulsive reaction forces can occur leading to an instantaneous change in the linear and angular momenta of each body. In this case, the aforementioned standard schemes completely fail since the numerical response may artificially generate energy when a contact occurs, see, e.g., [4]. In order to analyze such impact phenomena, nonsmooth time integration methods have been proposed in the literature and can be classified into two main groups, namely, event-driven schemes and time-stepping schemes. Event-driven schemes are based on an accurate event detection and the time step is adapted such that the end of the step coincides with an event. At this time instant, the event is solved with the help of an impact law. Such schemes are accurate for the free flight smooth motions and are especially suitable for small multi-body systems with a limited number of events, but they become inefficient if frequent transitions occur in a short time. Contrary to the event-driven schemes, time-stepping schemes do not adapt their time step size on events but only on some accuracy requirements if needed. Two main families of time-stepping schemes have been designed up to now: the Schatzman–Paoli scheme [5, 6] is based on a central difference scheme and the Moreau–Jean scheme [7, 8, 9] is based on a θ-method. Time-stepping methods have been proven to be convergent and robust, and are extensively applied as the solution to nonsmooth system models. In contrast to event-driven schemes, time-stepping schemes are expected to have order-one accuracy even in the smooth part of the motion. Thus, the accuracy is less satisfactory unless a very small step size is applied. However, they remain robust and efficient even for a large number of events. Some other strategies rely on an implicit detection of the impact times [10] and thus share properties of both event-driven and time–stepping schemes.

The algorithm proposed here for systems with rigid and/or flexible bodies is inspired from the Moreau–Jean time-stepping strategy. A fundamental property of the Moreau–Jean scheme is that the unilateral constraints are imposed at

velocity level. As recognized by many authors [1, 2, 3, 8, 9], this choice leads to interesting consistency and stability properties of a simulation algorithm for dynamic contact analysis. In the Moreau–Jean scheme, the unilateral constraints at position level are only used to support the decision to activate the constraint but they are not exactly satisfied. The consequence is that some penetration can be observed in the numerical solution, which may not be physically acceptable.

In order to prevent such penetration problems, this paper presents an algorithm which enforces the constraint not only at velocity level, so as to inherit good consistency and stability properties, but also at position level. For that purpose, the Gear-Gupta-Leimkuhler (GGL) approach, which was initially developed for the stabilization of index-3 differential-algebraic equations [11], is generalized to systems with unilateral constraints following a similar idea as in [12, 13]. After time discretization, the equations of motion involve one linear complementarity condition at position level and a second one at velocity level. However, the present work differs from [12, 13] by the choice of the Lagrange multipliers which appear in these complementarity conditions, as well as by the definition of an implicit algorithm which solves the dynamic equilibrium and the complementarity conditions in a monolithic way. At each time step, the problem is solved using a semi-smooth Newton process, which can also be interpreted as an active set method as discussed in [14, 15]. In order to guarantee the exact satisfaction of the unilateral constraints by the converged solution without any penetration, the active set is updated at each Newton iteration.

Another property of the Moreau–Jean algorithm is that the complete system is integrated in time using a method which is only first-order accurate. This appears as a disadvantage in applications where the nonsmooth phenomena are localized in some mechanical parts of the system, while other parts exhibit smooth motion. For example, the dynamic response of wind turbines systems is characterized by a coupling between smooth motions and vibrations of large structural components, such as the blades and the tower, and nonsmooth dynamic phenomena in the gearbox and in the transmission line. In this case, the description of vibration of phenomena in flexible bodies with a first-order method would require quite small time steps and would thus be inefficient.

In order to improve the accuracy in the smooth part of motion, Chen et al [16] observed that some terms in the equations of motion of a nonsmooth dynamic system, such as the elastic forces, are smooth and can be integrated in the time domain using a higher-order scheme, e.g., the generalized-α method. All impulsive terms are still treated using a fully implicit integration scheme to ensure consistency. This means that different integration formulae are used for the different contributions to the equations of motion. Globally, the order of convergence of the method is still limited to 1. Nevertheless, the advantage of this approach is that the numerical dissipation of the generalized-α method is significantly smaller than the numerical dissipation of the Euler implicit scheme, so that the energy behavior is strongly improved, especially for mechanical systems exhibiting both impacts and structural vibrations. In some sense, in this scheme,

the high numerical dissipation of the Euler implicit scheme is only acting locally at the contact region and during the contact time, and not on the full system during the whole trajectory. A similar idea is exploited here with two noticeable differences. Firstly, the smooth part of the reaction forces of bilateral constraints are now treated as smooth variables and integrated with a higher-order scheme, whereas in [16] the total reaction forces of bilateral constraints were integrated using the first-order scheme only. Secondly, the treatment of the unilateral and of the bilateral constraint both at velocity and position levels prevents from any drift-off phenomenon.

The paper is organized as follows. Section 2 formulates the equations of motion with unilateral and bilateral constraints as an equality of differential measures. Moreau's sweeping process is described in Section 3. Section 4 presents the splitting strategy which is used to isolate smooth terms in the equations of motion and the GGL formulation of the constraints is described in Section 5. The time discretization method is developed in Section 6 and its properties are studied in Section 7. A semi-smooth Newton algorithm which solves the discretized equations of motion at each time step is presented in Section 8. In order to perform some numerical comparisons, two algorithms from the literature, the Moreau–Jean θ-scheme and the smooth generalized-α method are briefly described in Section 9. Numerical examples with rigid impacts, accumulation phenomena, closed contact situations, and flexible contact conditions are studied in Section 10. Finally, the paper ends with some conclusions in Section 11.

2. Equations of motion

After spatial semi-discretization, the equations of motion of a flexible multibody system including bilateral and unilateral constraints can be expressed in the following form:

$$\begin{aligned} \dot{\mathbf{q}}^+ &= \mathbf{v}^+ && \text{(1a)} \\ \mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{v} - \mathbf{g}_{\mathbf{q}}^T\,\mathrm{d}\mathbf{i} &= \mathbf{f}(\mathbf{q},\mathbf{v},t)\,\mathrm{d}t && \text{(1b)} \\ \mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}) &= \mathbf{0} && \text{(1c)} \\ \mathbf{0} \le \mathbf{g}^{\mathcal{U}}(\mathbf{q}) &\perp \mathrm{d}\mathbf{i}^{\mathcal{U}} \ge \mathbf{0} && \text{(1d)} \end{aligned}$$

where

- $\mathbf{q}$ is the vector of coordinates, e.g., absolute nodal coordinates;
- $\dot{\mathbf{q}}^+(t) = \lim_{\tau\to t,\tau>t} \dot{\mathbf{q}}(\tau)$ and $\mathbf{v}^+(t) = \lim_{\tau\to t,\tau>t} \mathbf{v}(\tau)$ are the right limits of the velocity, which are functions of bounded variations;
- $\mathbf{f} = \mathbf{f}^{ext}(t) - \mathbf{f}^{damp}(\mathbf{q},\mathbf{v}) - \mathbf{f}^{int}(\mathbf{q})$ collects the external, damping and internal forces;
- $\mathbf{M}$ is the mass matrix which may, in general, depend on the coordinates;

- $\mathrm{d}\mathbf{v}$ is the differential measure associated with the velocity $\mathbf{v}$ assumed to be of bounded variations;
- t is time, and $\mathrm{d}t$ is the corresponding standard Lebesgue measure;
- $\mathbf{g}$ is the combined set of bilateral and unilateral constraints, and $\mathbf{g}_{\mathbf{q}}(\mathbf{q})$ is the corresponding matrix of constraint gradients;
- $\mathrm{d}\mathbf{i}$ is the impulse measure of the contact reactions and of the bilateral forces;
- $\mathcal{U}$ denotes the set of indices of the unilateral constraints, $\overline{\mathcal{U}}$ is its complementarity set, i.e., the set of bilateral constraints, $\mathcal{C} = \mathcal{U} \cup \overline{\mathcal{U}}$ is the full set of constraints, and we have

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}^{\mathcal{U}} \\ \mathbf{g}^{\overline{\mathcal{U}}} \end{bmatrix}, \qquad \mathrm{d}\mathbf{i} = \begin{bmatrix} \mathrm{d}\mathbf{i}^{\mathcal{U}} \\ \mathrm{d}\mathbf{i}^{\overline{\mathcal{U}}} \end{bmatrix} \tag{2}$$

For the sake of notation simplicity, the convention $\mathbf{v}(t) = \mathbf{v}^+(t)$ and $\dot{\mathbf{q}}(t) = \dot{\mathbf{q}}^+(t)$ shall be used in the remaining part of the paper.

Since the motion might be nonsmooth, jumps in the velocity and impacts are expected yielding the following decomposition of the measures, neglecting singular measures:

$$\mathrm{d}\mathbf{v} = \dot{\mathbf{v}}\,\mathrm{d}t + \sum_i (\mathbf{v}(t_i) - \mathbf{v}^-(t_i))\,\delta_{t_i} \tag{3}$$

$$\mathrm{d}\mathbf{i} = \boldsymbol{\lambda}\,\mathrm{d}t + \sum_i \mathbf{p}_i\,\delta_{t_i} \tag{4}$$

where $\boldsymbol{\lambda}$ is the vector of nonimpulsive Lagrange multipliers associated with the Lebesgue measurable constraint forces; $\mathbf{v}^-(t) = \lim_{\tau \to t, \tau < t} \mathbf{v}(\tau)$; $(\mathbf{v}(t_i) - \mathbf{v}^-(t_i))$ is the jump in velocity at the instant t_i, δ_{t_i} is the Dirac delta supported at t_i, and $\mathbf{p}_i$ is the impulse producing the jump at the instant t_i.

Equation (1d) is formulated as a complementarity condition, i.e., $(\mathbf{g}^{\mathcal{U}}(\mathbf{q}))^T\,\mathrm{d}\mathbf{i}^{\mathcal{U}} = 0$ with both $\mathbf{g}^{\mathcal{U}}(\mathbf{q})$ and $\mathrm{d}\mathbf{i}^{\mathcal{U}}$ being non-negative. To complete the model, an impact law is needed. The Newton impact law defines the normal velocity jump in case of an impact for the constraint $j \in \mathcal{U}$ as $g^j_{\mathbf{q}}\,\mathbf{v}(t) = -e^j\,g^j_{\mathbf{q}}\,\mathbf{v}^-(t)$, where $e^j \in [0,1]$ is the coefficient of restitution, and $\mathbf{v}^-(t)$ and $\mathbf{v}(t)$ denote the velocity before and after the impact, respectively. Therefore, the contact condition is expressed at velocity level as [9]

$$\text{if } g^j(\mathbf{q}(t_i)) \leq 0 \text{ then } 0 \leq g^j_{\mathbf{q}}\,\mathbf{v}(t_i) + e^j\,g^j_{\mathbf{q}}\,\mathbf{v}^-(t_i) \perp p^j_i \geq 0, \quad \forall j \in \mathcal{U} \tag{5}$$

The present formalism is developed for the analysis of contact conditions between rigid or flexible bodies. For rigid bodies, the coefficient of restitution defines the amount of energy dissipated during an impact. For flexible bodies, the physical meaning of a coefficient of restitution is not clear. The spatial discretization of a

flexible body using the finite element method leads to a finite dimensional system with finite masses. An impact law with a coefficient of restitution is thus needed to describe contact conditions. In practice, for flexible bodies, a value $e^j = 0$ may be used so that the condition $g^j_{\mathbf{q}}\,\mathbf{v}(t_i) = 0$ is imposed when the constraint is active.

Inserting Eqs. (3-4) in (1a-1d) yields the standard equations of motion for almost every time t:

$$\begin{aligned} \mathbf{M}(\mathbf{q})\dot{\mathbf{v}} - \mathbf{g}_{\mathbf{q}}^T\,\boldsymbol{\lambda} &= \mathbf{f}(\mathbf{q},\mathbf{v},t) && \text{(6a)}\\ \mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}) &= \mathbf{0} && \text{(6b)}\\ \mathbf{0} \le \mathbf{g}^{\mathcal{U}}(\mathbf{q}) &\perp \boldsymbol{\lambda}^{\mathcal{U}} \ge \mathbf{0} && \text{(6c)} \end{aligned}$$

and the impact equations at time t_i

$$\begin{aligned} \mathbf{M}(\mathbf{q})\,(\mathbf{v}(t_i) - \mathbf{v}^-(t_i)) - \mathbf{g}_{\mathbf{q}}^T\,\mathbf{p}_i &= \mathbf{0} && \text{(6d)}\\ \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\,\mathbf{v}(t_i) &= \mathbf{0} && \text{(6e)}\\ \text{if } g^j(\mathbf{q}(t_i)) \le 0 \text{ then } 0 \le g^j_{\mathbf{q}}\,\mathbf{v}(t_i) + e^j\,g^j_{\mathbf{q}}\,\mathbf{v}^-(t_i) &\perp p_i^j \ge \mathbf{0}, \quad \forall j \in \mathcal{U} && \text{(6f)} \end{aligned}$$

3. Moreau's sweeping process

The Moreau–Jean method is based on a reformulation of the equations of motion with the constraints at velocity level. For unilateral constraints, the constraint at velocity level is provided by the Newton's impact law. This leads to the following set of equations

$$\begin{aligned} \dot{\mathbf{q}} &= \mathbf{v} && \text{(7a)}\\ \mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{v} - \mathbf{g}_{\mathbf{q}}^T\,\mathrm{d}\mathbf{i} &= \mathbf{f}(\mathbf{q},\mathbf{v},t)\,\mathrm{d}t && \text{(7b)}\\ \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\,\mathbf{v} &= \mathbf{0} && \text{(7c)}\\ \text{if } g^j(\mathbf{q}) \le 0 \text{ then } 0 \le g^j_{\mathbf{q}}\,\mathbf{v} + e^j\,g^j_{\mathbf{q}}\,\mathbf{v}^- &\perp \mathrm{d}i^j \ge 0, \quad \forall\, j \in \mathcal{U} && \text{(7d)} \end{aligned}$$

As a consequence of Moreau's viability Lemma [17], the exact solution of these equations also satisfies the complementarity condition at position level in Eq. (1d).

If all unilateral constraints are closed, the original system of equations (1) with the constraints at position level is an index-3 DAE, while the system (7) with the constraints at velocity level is an index-2 DAE. From a numerical point of view, the difficulties encountered when solving an index-2 DAE are less severe than when solving the original index-3 DAE [18]. Therefore, the reformulation of the equations of motion in (7) can be interpreted as an index reduction process for systems with unilateral constraints. A time discretization of Eq. (7) leads to a discrete system in which the constraints are imposed at velocity level. This is an important property to guarantee the stability and the consistency of the numerical solution for nonsmooth dynamic systems with unilateral constraints and impacts. The consequence is that the constraints at velocity level are exactly satisfied but,

due to the presence of round-off and truncation errors, the constraints at position level are only approximately satisfied by the numerical solution. This leads to a drift phenomenon which may be characterized by a progressive increase in the constraint violation at position level during the simulation. Starting from the formulation in Eq. (7), this paper aims at developing a consistent approach which guarantees the exact satisfaction of the constraints both at velocity and position levels, in order to completely eliminate any drift problem.

4. Splitting strategy

Following a similar approach as in [16], a splitting of the variables and of the equations of motion is proposed to isolate impulsive terms, which are integrated with only first order accuracy, from the other contributions, which are assumed to be sufficiently smooth to be integrated using a higher order method. In this paper, the splitting will also be exploited to derive the GGL formulation presented in the next section.

For a given time step $(t_n, t_{n+1}]$, the position $\tilde{\mathbf{q}}(t)$, velocity $\tilde{\mathbf{v}}(t)$ and Lagrange multiplier $\tilde{\boldsymbol{\lambda}}(t)$ are defined as the solution of the following initial value problem coupled with the variables $\mathbf{q}(t)$ and $\mathbf{v}(t)$

$$\begin{aligned} \dot{\tilde{\mathbf{q}}} &= \tilde{\mathbf{v}} && \text{(8a)} \\ \mathbf{M}(\mathbf{q})\,\dot{\tilde{\mathbf{v}}} - \mathbf{g}_{\mathbf{q}}^{T}(\mathbf{q})\,\tilde{\boldsymbol{\lambda}} &= \mathbf{f}(\mathbf{q}, \mathbf{v}, t) && \text{(8b)} \\ \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}(\mathbf{q})\,\tilde{\mathbf{v}} &= \mathbf{0} && \text{(8c)} \\ \tilde{\boldsymbol{\lambda}}^{\mathcal{U}} &= \mathbf{0} && \text{(8d)} \end{aligned}$$

with the initial value $\tilde{\mathbf{v}}(t_n) = \mathbf{v}(t_n)$, $\tilde{\mathbf{q}}(t_n) = \mathbf{q}(t_n)$. If there is no unilateral constraint, one observes that $\tilde{\mathbf{v}}$ exactly satisfies the equations of motion (6a) with the bilateral constraints at velocity level, so that $\tilde{\mathbf{v}} = \mathbf{v}$ and $\tilde{\boldsymbol{\lambda}} = \boldsymbol{\lambda}$. For unilaterally constrained systems, Eq. (8) thus represent the dynamics of the system if the unilateral constraints and the contact forces are partially ignored on the interval $(t_n, t_{n+1}]$. Indeed, the unilateral constraints still have an indirect influence on the smooth solution $\tilde{\mathbf{q}}(t)$, $\tilde{\mathbf{v}}(t)$ because Eq. (8) is still coupled with the variables $\mathbf{q}(t)$, $\mathbf{v}(t)$. In the literature, $\tilde{\mathbf{v}}$ is sometimes called the free velocity, i.e., the velocity of the system when the reaction forces of the unilateral constraints are null [19]. If $\mathbf{f}(\mathbf{q}(t), \mathbf{v}(t), t)$ is a sufficiently smooth function of time and if the matrix

$$\begin{bmatrix} \mathbf{M} & \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T} \\ \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}} & \mathbf{0} \end{bmatrix} \tag{9}$$

is nonsingular, Eq. (8) is a well-defined and smooth system of index-2 differential-algebraic equations. Let us remark that the mass matrix itself can be singular.

Then, the variable $\mathrm{d}\mathbf{w}$ is defined by the equation

$$\mathrm{d}\mathbf{w} = \mathrm{d}\mathbf{v} - \dot{\tilde{\mathbf{v}}}\,\mathrm{d}t \tag{10}$$

The motivation for this splitting is to express $\mathrm{d}\mathbf{v}$ as the sum of a smooth contribution which is represented by the variable $\dot{\tilde{\mathbf{v}}}$ and of some impulsive contributions which are all collected by the variable $\mathrm{d}\mathbf{w}$. The variable $\mathrm{d}\mathbf{w}$ includes all discontinuous contributions to the velocity, which result from the various impacts that occur during the time step. Based on this idea, $\tilde{\mathbf{v}}$ can be evaluated by time integration using a second-order method while, to ensure consistency, $\mathrm{d}\mathbf{w}$ is evaluated using the Moreau–Jean scheme and the Euler implicit method. However, second-order accuracy cannot always be guaranteed for $\tilde{\mathbf{v}}$ since $\mathbf{f}(\mathbf{q}(t), \mathbf{v}(t), t)$ can be a discontinuous function of time whenever a velocity jump occurs. An elimination of $\mathrm{d}\mathbf{v}$ and $\dot{\tilde{\mathbf{v}}}$ from Eqs. (7b,8b,10) yields

$$\mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{w} - \mathbf{g}_{\mathbf{q}}^{T}\,(\mathrm{d}\mathbf{i} - \tilde{\boldsymbol{\lambda}}\,\mathrm{d}t) = \mathbf{0} \tag{11}$$

In terms of these variables, the equations of motion become

$$\begin{aligned}
\dot{\mathbf{q}} &= \mathbf{v} && \text{(12a)}\\
\mathrm{d}\mathbf{v} &= \mathrm{d}\mathbf{w} + \dot{\tilde{\mathbf{v}}}\,\mathrm{d}t && \text{(12b)}\\
\mathbf{M}(\mathbf{q})\,\dot{\tilde{\mathbf{v}}} - \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T}\,\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}} &= \mathbf{f}(\mathbf{q}, \mathbf{v}, t) && \text{(12c)}\\
\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\,\tilde{\mathbf{v}} &= \mathbf{0} && \text{(12d)}\\
\tilde{\boldsymbol{\lambda}}^{\mathcal{U}} &= \mathbf{0} && \text{(12e)}\\
\mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{w} - \mathbf{g}_{\mathbf{q}}^{T}\,(\mathrm{d}\mathbf{i} - \tilde{\boldsymbol{\lambda}}\,\mathrm{d}t) &= \mathbf{0} && \text{(12f)}\\
\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\mathbf{v} &= \mathbf{0} && \text{(12g)}\\
\text{if } g^j(\mathbf{q}) \le 0 \text{ then } 0 \le g_{\mathbf{q}}^j\,\mathbf{v} + e^j\,g_{\mathbf{q}}^j\,\mathbf{v}^- &\perp \mathrm{d}i^j \ge 0, \quad \forall j \in \mathcal{U} && \text{(12h)}
\end{aligned}$$

The set of 3 equations (12b, 12c, 12f) for the three variables $\mathrm{d}\mathbf{v}$, $\dot{\tilde{\mathbf{v}}}$ and $\mathrm{d}\mathbf{w}$ is strictly equivalent to the single Eq. (7b) for the variable $\mathrm{d}\mathbf{v}$.

5. Gear-Gupta-Leimkuhler formulation

The equations of motion (12) involve the bilateral and unilateral constraints at velocity level only. As mentioned in Section 3, the exact solution would also satisfy the constraints at position level

$$\begin{aligned}
\mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}) &= \mathbf{0} && \text{(13a)}\\
\mathbf{g}^{\mathcal{U}}(\mathbf{q}) &\ge \mathbf{0} && \text{(13b)}
\end{aligned}$$

However, a time integration algorithm based on Eq. (12) that enforces the constraints at velocity level only will not satisfy the constraints at position level because of the drift phenomenon. In this section, we propose to include the constraints at velocity and position levels in a unique set of equations following a method similar to that proposed by Gear, Gupta and Leimkuhler [11]. The advantage of this strategy is that the time discretization procedure will lead to an

algorithm which simultaneously satisfies the constraints at position and velocity levels.

The idea is to supplement Eq. (12) with the constraints at position level. Equation (12a) is multiplied by the mass matrix

$$\mathbf{M}(\mathbf{q})\dot{\mathbf{q}} = \mathbf{M}(\mathbf{q})\mathbf{v} \tag{14}$$

and an additional Lagrange multiplier $\boldsymbol{\mu}$ is introduced in order to accommodate for this duplicated set of position and velocity constraints and to have a well-posed problem. Therefore, Eq. (14) is modified as

$$\mathbf{M}(\mathbf{q})\,\dot{\mathbf{q}} - \mathbf{g}_{\mathbf{q}}^T\,\boldsymbol{\mu} = \mathbf{M}(\mathbf{q})\,\mathbf{v} \tag{15}$$

where it is expected that $\boldsymbol{\mu}$ vanishes identically for the exact solution as a consequence of the redundancy between the constraints at position and velocity levels. As a consequence, the variables $\dot{\mathbf{q}}$ and $\mathbf{v}$ are not formally equivalent, the variable $\dot{\mathbf{q}}$ being related with the position variable $\mathbf{q}$ through the equation

$$\mathbf{q}(t) = \mathbf{q}(t_n) + \int_{(t_n,t]} \dot{\mathbf{q}}(\tau)\mathrm{d}\tau \tag{16}$$

Following [12], the equilibrium equations are then obtained after replacing Eq. (12) by

$$\begin{aligned}
\mathbf{M}(\mathbf{q})\,\dot{\mathbf{q}} - \mathbf{g}_{\mathbf{q}}^T\,\boldsymbol{\mu} &= \mathbf{M}(\mathbf{q})\,\mathbf{v} && \text{(17a)}\\
\mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}) &= \mathbf{0} && \text{(17b)}\\
\mathbf{0} \le \mathbf{g}^{\mathcal{U}}(\mathbf{q}) &\perp \boldsymbol{\mu}^{\mathcal{U}} \ge \mathbf{0} && \text{(17c)}\\
\mathrm{d}\mathbf{v} &= \mathrm{d}\mathbf{w} + \dot{\tilde{\mathbf{v}}}\,\mathrm{d}t && \text{(17d)}\\
\mathbf{M}(\mathbf{q})\,\dot{\tilde{\mathbf{v}}} - \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T}\,\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}} &= \mathbf{f}(\mathbf{q},\mathbf{v},t) && \text{(17e)}\\
\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\,\tilde{\mathbf{v}} &= \mathbf{0} && \text{(17f)}\\
\tilde{\boldsymbol{\lambda}}^{\mathcal{U}} &= \mathbf{0} && \text{(17g)}\\
\mathbf{M}(\mathbf{q})\,\mathrm{d}\mathbf{w} - \mathbf{g}_{\mathbf{q}}^T\,(\mathrm{d}\mathbf{i} - \tilde{\boldsymbol{\lambda}}\,\mathrm{d}t) &= \mathbf{0} && \text{(17h)}\\
\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\mathbf{v} &= \mathbf{0} && \text{(17i)}\\
\text{if } g^j(\mathbf{q}) \le 0 \text{ then } 0 \le g_{\mathbf{q}}^j\,\mathbf{v} + e^j\,g_{\mathbf{q}}^j\,\mathbf{v}^- &\perp \mathrm{d}i^j \ge 0, \quad \forall j \in \mathcal{U} && \text{(17j)}
\end{aligned}$$

By construction, any solution of Eq. (12) also satisfies Eq. (17) with $\boldsymbol{\mu} = \mathbf{0}$. Let us also observe that this formulation is still valid when the mass matrix is singular, provided that the matrix in Eq. (9) is nonsingular.

If all unilateral constraints are closed, the GGL form of the equations of motion is an index-2 DAE. In this special case, the formalism presented here boils down to the DAE formalism proposed by Gear, Gupta and Leimkuhler [11]. An important difference compared to the index-2 DAE in Eq. (7) is that the constraints are now imposed both at velocity and position levels.

6. Time discretization

By construction, the variables $\tilde{\mathbf{q}}(t)$, $\tilde{\mathbf{v}}(t)$ and $\tilde{\boldsymbol{\lambda}}(t)$ only capture a sufficiently smooth part of the motion during the time step $(t_n, t_{n+1}]$ and they shall be computed using a second-order time integration scheme. In order to obtain the complete time integration scheme, discrete nonsmooth corrections are now defined so that the total motion satisfies a discrete form of the bilateral constraints in Eqs. (17b,17i) as well as a discrete form of the complementarity conditions in Eqs. (17c,17j), which represent the unilateral constraints.

6.1. Discrete nonsmooth variables

The velocity jump and position correction variables $\mathbf{W}(t_n, t)$ and $\mathbf{U}(t_n; t)$ are defined as

$$\mathbf{W}(t_n; t) = \int_{(t_n,t]} \mathrm{d}\mathbf{w} \tag{18a}$$

$$\mathbf{U}(t_n; t) = \int_{t_n}^{t} (\dot{\mathbf{q}} - \tilde{\mathbf{v}}) \mathrm{d}t \tag{18b}$$

Since $\mathbf{v}(t_n) = \tilde{\mathbf{v}}(t_n)$ and $\mathbf{q}(t_n) = \tilde{\mathbf{q}}(t_n)$, we get

$$\mathbf{W}(t_n; t) = \mathbf{v}(t) - \tilde{\mathbf{v}}(t) \tag{19a}$$

$$\mathbf{U}(t_n; t) = \mathbf{q}(t) - \tilde{\mathbf{q}}(t) \tag{19b}$$

as well as $\mathbf{W}(t_n; t_n) = \mathbf{0}$ and $\mathbf{U}(t_n; t_n) = \mathbf{0}$. The multipliers $\boldsymbol{\Lambda}(t_n; t)$ and $\boldsymbol{\nu}(t_n; t)$ are defined as

$$\boldsymbol{\Lambda}(t_n; t) = \int_{(t_n,t]} (\mathrm{d}\mathbf{i} - \tilde{\boldsymbol{\lambda}}(\tau)\, \mathrm{d}\tau) \tag{20a}$$

$$\boldsymbol{\nu}(t_n; t) = \int_{t_n}^{t} (\boldsymbol{\mu}(\tau) + \boldsymbol{\Lambda}(t_n; \tau))\, \mathrm{d}\tau \tag{20b}$$

with $\boldsymbol{\Lambda}(t_n; t_n) = \boldsymbol{\nu}(t_n; t_n) = \mathbf{0}$.

The discrete approximation of Eqs. (17a) and (17h) is based on the following result.

Theorem 1. *The velocity jump and position correction variables, which are caused by the unilateral constraints, satisfy*

$$\mathbf{M}(\mathbf{q}(t_{n+1}))\, \mathbf{W}(t_n; t_{n+1}) - \mathbf{g}_{\mathbf{q}}^T(\mathbf{q}(t_{n+1}))\, \boldsymbol{\Lambda}(t_n; t_{n+1}) = \mathcal{O}(h) \tag{21a}$$

$$\mathbf{M}(\mathbf{q}(t_{n+1}))\, \mathbf{U}(t_n; t_{n+1}) - \mathbf{g}_{\mathbf{q}}^T(\mathbf{q}(t_{n+1}))\, \boldsymbol{\nu}(t_n; t_{n+1}) = \mathcal{O}(h^2) \tag{21b}$$

in which $h = t_{n+1} - t_n$ is the step size.

Proof. The proof is obtained by integrating Eqs. (17a) and (17h) over the time step $(t_n, t_{n+1}]$.

Firstly, we study the integrated form of Eq. (17h). Since $\mathbf{q}(t)$ is continuous, we have $\mathbf{M}(\mathbf{q}(\tau)) = \mathbf{M}(\mathbf{q}(t)) + \mathcal{O}(h)$ and $\mathbf{g}_\mathbf{q}^T(\mathbf{q}(\tau)) = \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) + \mathcal{O}(h)$, $\forall \tau \in (t_n, t]$. Then, Eq. (21a) is obtained by observing that, $\forall t \in (t_n, t_{n+1}]$,

$$\begin{aligned} \mathbf{0} &= \int_{(t_n,t]} \Big(\mathbf{M}(\mathbf{q}) \mathrm{d}\mathbf{w} - \mathbf{g}_\mathbf{q}^T \left(\mathrm{d}\mathbf{i} - \tilde{\boldsymbol{\lambda}} \, \mathrm{d}\tau \right) \Big) \\ &= \mathbf{M}(\mathbf{q}(t)) \, \mathbf{W}(t_n; t) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) \, \boldsymbol{\Lambda}(t_n; t) + \mathcal{O}(h) \end{aligned} \tag{22}$$

Secondly, we study the integrated form of Eq. (17a). Using $\mathbf{M}(\mathbf{q}(t)) = \mathbf{M}(\mathbf{q}(t_{n+1})) + \mathcal{O}(h)$, $\mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) = \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t_{n+1})) + \mathcal{O}(h)$, $\forall t \in (t_n, t_{n+1}]$, and $\mathbf{q}_n = \tilde{\mathbf{q}}_n$, the integrated form of Eq. (17a) is obtained as

$$\begin{aligned} \mathbf{0} &= \int_{t_n}^{t_{n+1}} \left(\mathbf{M}(\mathbf{q}(t))(\dot{\mathbf{q}}(t) - \mathbf{v}(t)) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) \, \boldsymbol{\mu}(t) \right) \mathrm{d}t \\ &= \int_{t_n}^{t_{n+1}} \left(\mathbf{M}(\mathbf{q}(t))(\dot{\mathbf{q}}(t) - \tilde{\mathbf{v}}(t) - \mathbf{W}(t_n; t)) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) \, \boldsymbol{\mu}(t) \right) \mathrm{d}t \\ &= \int_{t_n}^{t_{n+1}} \left(\mathbf{M}(\mathbf{q}(t))(\dot{\mathbf{q}}(t) - \tilde{\mathbf{v}}(t)) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t)) \, (\boldsymbol{\Lambda}(t_n; t) + \boldsymbol{\mu}(t)) + \mathcal{O}(h) \right) \mathrm{d}t \\ &= \mathbf{M}(\mathbf{q}(t_{n+1})) \, (\mathbf{q}(t_{n+1}) - \mathbf{q}(t_n) - (\tilde{\mathbf{q}}(t_{n+1}) - \tilde{\mathbf{q}}(t_n))) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t_{n+1})) \, \boldsymbol{\nu}(t_n; t_{n+1}) + \mathcal{O}(h^2) \\ &= \mathbf{M}(\mathbf{q}(t_{n+1})) \mathbf{U}(t_n; t_{n+1}) - \mathbf{g}_\mathbf{q}^T(\mathbf{q}(t_{n+1})) \, \boldsymbol{\nu}(t_n; t_{n+1}) + \mathcal{O}(h^2) \end{aligned}$$

where Eqs. (19a), (22), (20b) and (19b) have been successively used. □

Let us remark that, if impacts occur and $h \to 0$, $\mathbf{W}(t_n; t_{n+1}) = \mathcal{O}(1)$ and $\boldsymbol{\Lambda}(t_n; t_{n+1}) = \mathcal{O}(1)$ since the velocities may exhibit finite jumps, whereas $\mathbf{U}(t_n; t_{n+1}) = \mathcal{O}(h)$ and $\boldsymbol{\nu}(t_n; t_{n+1}) = \mathcal{O}(h)$ since the positions remain continuous. The position corrections eliminate the drift of the constraints, and the velocity jumps are approximated over the time step.

6.2. Discrete complementarity conditions

The discrete complementarity conditions at velocity and position levels can be obtained from the inequalities $\mathrm{d}\mathbf{i}^\mathcal{U} \geq \mathbf{0}$ in Eq. (17j), and $\boldsymbol{\mu}^\mathcal{U} \geq \mathbf{0}$ in Eq. (17c).

At velocity level, Eqs. (17g) and (20a) imply that

$$\int_{(t_n,t_{n+1}]} \mathrm{d}\mathbf{i}^\mathcal{U} = \boldsymbol{\Lambda}^\mathcal{U}(t_n; t_{n+1}) \tag{23}$$

so that, following a similar idea as in the Moreau–Jean method, the complementarity condition at velocity level is expressed as

$$\text{if } g^j(\mathbf{q}_{n+1}^*) \leq 0 \text{ then } 0 \leq g_{\mathbf{q},n+1}^j \, \mathbf{v}_{n+1} + e^j \, g_{\mathbf{q},n}^j \, \mathbf{v}_n \perp \Lambda_{n+1}^j \geq 0, \quad \forall j \in \mathcal{U} \tag{24}$$

At position level, a fully implicit complementarity expression between the collocated value of the constraint $\mathbf{g}^{\mathcal{U}}(\mathbf{q})$ at time t_{n+1} and the multiplier $\boldsymbol{\mu}^{\mathcal{U}}(t)$ at time t_{n+1} would lead to

$$\mathbf{0} \leq \mathbf{g}^{\mathcal{U}}(\mathbf{q}_{n+1}) \perp \boldsymbol{\mu}^{\mathcal{U}}_{n+1} \geq \mathbf{0} \tag{25}$$

This relation is written in terms of the impulse $\boldsymbol{\mu}^{\mathcal{U}}$, which is not directly accessible in the algorithm (integrated values along the time step are instead used, i.e., $\boldsymbol{\nu}^{\mathcal{U}}(t_n; t_{n+1})$ and $\boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t_{n+1})$).

Starting from Eq. (20b), we can get by using the θ-method

$$\boldsymbol{\nu}^{\mathcal{U}}(t_n; t_{n+1}) = h\left(\theta\left(\boldsymbol{\mu}^{\mathcal{U}}(t_{n+1}) + \boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t_{n+1})\right) + (1-\theta)\boldsymbol{\mu}^{\mathcal{U}}(t_n)\right) + \mathcal{O}(h). \tag{26}$$

This result can be obtained from Lemma 1 in [20], assuming that $\boldsymbol{\mu}^{\mathcal{U}}(t) + \boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t)$ is a bounded variation function. The discontinuous behaviour of $\boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t)$ in the interval $t \in (t_n, t_{n+1}]$ makes it difficult to get a better numerical approximation, because the contact time instant is not known. Only in the special case where no impact occurs in the interval $(t_n, t_{n+1}]$, the variable $\boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t)$ is continuous and the local truncation error in Eq. (26) drops to $\mathcal{O}(h^2)$.

For instance, a fully implicit Euler approximation with $\theta = 1$ leads to

$$h\boldsymbol{\mu}^{\mathcal{U}}(t_{n+1}) \simeq \boldsymbol{\nu}^{\mathcal{U}}(t_n; t_{n+1}) - h\boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t_{n+1}) \tag{27}$$

However, both terms $\boldsymbol{\nu}^{\mathcal{U}}(t_n; t_{n+1})$ and $h\boldsymbol{\Lambda}^{\mathcal{U}}(t_n; t_{n+1})$ are $\mathcal{O}(h)$, so that the relative error on this estimate is at best $\mathcal{O}(1)$, i.e., it is dominated by numerical artifacts. Experiments using this approximation in Eq. (25) evidenced spurious chattering phenomena, as shown in Section 10.1. For this reason, an alternative approach is proposed.

Note that, if contact phenomena occur during the time step $(t_n; t_{n+1}]$, then $\mu^j(t_n; t_{n+1}) \geq 0$, $\Lambda^j(t_n; t_{n+1}) > 0$ and therefore $\nu^j(t_n; t_{n+1}) > 0$. Based on this observation, we make the assumption that the numerical solution should satisfy the complementarity condition

$$\mathbf{0} \leq \mathbf{g}^{\mathcal{U}}(\mathbf{q}_{n+1}) \perp \boldsymbol{\nu}^{\mathcal{U}}_{n+1} \geq \mathbf{0} \tag{28}$$

From Eq. (21b), the situation $\nu^j(t_n; t_{n+1}) > 0$ implies that $\mathbf{U}(t_n, t_{n+1}) \neq \mathbf{0}$, i.e., that the trajectory departs from the smooth trajectory and contact phenomena have occurred needing a position correction, imposed by the constraint $g^j(\mathbf{q}_{n+1}) = 0$. Conversely, if $g^j(\mathbf{q}_{n+1}) > 0$, then there is no impact at time t_{n+1}, no contact interaction has occurred and the condition $\nu^j_{n+1} = 0$ needs to be satisfied. This assumption does not affect the order of accuracy of the algorithm since it specifies the influence of an $\mathcal{O}(h)$ correction to the position variable, while $\mathcal{O}(h)$ local errors are anyway accepted when impacts occur. However, this assumption is essential because it significantly affects the qualitative behaviour of the numerical solution. It will be shown below that it results in a robust and chattering-free numerical scheme.

6.3. Time stepping scheme

The complete time integration scheme is obtained by combining a first-order approximation of the nonsmooth variables with a one-step and second-order time integration scheme for the smooth variables. In this work, the generalized-α method is used for the smooth part, however, other DAE time integration schemes could also be considered. According to [21], the generalized-α method combines the equations of motion at time step $n+1$ with some difference equations. This hybrid time integration scheme permitting to advance the solution at each step is formulated as follows

$$\begin{aligned}
\mathbf{M}(\mathbf{q}_{n+1})\dot{\tilde{\mathbf{v}}}_{n+1} - \mathbf{f}(\mathbf{q}_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) - \mathbf{g}^{\overline{\mathcal{U}},T}_{\mathbf{q},n+1}\,\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}}_{n+1} &= \mathbf{0} && \text{(29a)}\\
\mathbf{g}^{\overline{\mathcal{U}}}_{\mathbf{q},n+1}\,\tilde{\mathbf{v}}_{n+1} &= \mathbf{0} && \text{(29b)}\\
\mathbf{M}(\mathbf{q}_{n+1})\mathbf{U}_{n+1} - \mathbf{g}^{T}_{\mathbf{q},n+1}\,\boldsymbol{\nu}_{n+1} &= \mathbf{0} && \text{(29c)}\\
\mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}_{n+1}) &= \mathbf{0} && \text{(29d)}\\
\mathbf{0} \le \mathbf{g}^{\mathcal{U}}(\mathbf{q}_{n+1}) \perp \boldsymbol{\nu}^{\mathcal{U}}_{n+1} &\ge \mathbf{0} && \text{(29e)}\\
\mathbf{M}(\mathbf{q}_{n+1})\mathbf{W}_{n+1} - \mathbf{g}^{T}_{\mathbf{q},n+1}\boldsymbol{\Lambda}_{n+1} &= \mathbf{0} && \text{(29f)}\\
\mathbf{g}^{\overline{\mathcal{U}}}_{\mathbf{q},n+1}\mathbf{v}_{n+1} &= \mathbf{0} && \text{(29g)}\\
\text{if } g^j(\mathbf{q}^*_{n+1}) \le 0 \text{ then } 0 \le g^j_{\mathbf{q},n+1}\,\mathbf{v}_{n+1} + e^j\, g^j_{\mathbf{q},n}\,\mathbf{v}_n \perp \Lambda^j_{n+1} &\ge 0, \quad \forall j \in \mathcal{U} && \text{(29h)}
\end{aligned}$$

with the notation $\mathbf{g}^{T}_{\mathbf{q},n+1} = \mathbf{g}^{T}_{\mathbf{q}}(\mathbf{q}_{n+1})$ and the difference equations

$$\begin{aligned}
\tilde{\mathbf{q}}_{n+1} &= \mathbf{q}_n + h\mathbf{v}_n + h^2(0.5-\beta)\mathbf{a}_n + h^2\beta\mathbf{a}_{n+1} && \text{(30a)}\\
\mathbf{q}_{n+1} &= \tilde{\mathbf{q}}_{n+1} + \mathbf{U}_{n+1} && \text{(30b)}\\
\tilde{\mathbf{v}}_{n+1} &= \mathbf{v}_n + h(1-\gamma)\mathbf{a}_n + h\gamma\mathbf{a}_{n+1} && \text{(30c)}\\
\mathbf{v}_{n+1} &= \tilde{\mathbf{v}}_{n+1} + \mathbf{W}_{n+1} && \text{(30d)}\\
(1-\alpha_m)\mathbf{a}_{n+1} + \alpha_m\mathbf{a}_n &= (1-\alpha_f)\dot{\tilde{\mathbf{v}}}_{n+1} + \alpha_f\dot{\tilde{\mathbf{v}}}_n && \text{(30e)}
\end{aligned}$$

The numerical coefficients γ, β, α_m, and α_f can be selected from a desired value of the spectral radius at the infinity $\rho_\infty \in [0,1]$, see [22]

$$\alpha_m = \frac{2\rho_\infty - 1}{\rho_\infty + 1}, \quad \alpha_f = \frac{\rho_\infty}{\rho_\infty + 1}, \quad \gamma = 0.5 + \alpha_f - \alpha_m, \quad \beta = 0.25(\gamma + 0.5)^2 \quad (31)$$

The following remarks can be formulated.

1. The variable $\mathbf{a}_{n+1}$ is a pseudo-acceleration variable which is initialized as $\mathbf{a}_0 = \dot{\tilde{\mathbf{v}}}_0$. This variable can be interpreted as an approximation of the true acceleration $\dot{\tilde{\mathbf{v}}}$ at a shifted time, i.e., $\mathbf{a}_{n+1} \simeq \dot{\tilde{\mathbf{v}}}(t_{n+1} + (\alpha_m - \alpha_f)h)$. Ref. [21] includes a more detailed discussion about the generalized-α method for the simulation of constrained mechanical systems.

2. As shown in [23], the generalized-α method is mathematically equivalent to a two-step method for the velocity variable $\tilde{\mathbf{v}}$ and to a three-step method for the position variable $\tilde{\mathbf{q}}$. Indeed, the value of the acceleration-like variable $\mathbf{a}_n$ depends on the history of $\dot{\tilde{\mathbf{v}}}$. Therefore, the time integration of $\dot{\tilde{\mathbf{v}}}$ over the time step $(t_n, t_{n+1}]$ implicitly exploits some values of $\dot{\tilde{\mathbf{v}}}(t)$ for $t < t_n$. When a jump in the velocity $\mathbf{v}$ occurs, the forces $\mathbf{f}(\mathbf{q}, \mathbf{v}, t)$ in Eq. (8b) are also discontinuous and so is the variable $\dot{\tilde{\mathbf{v}}}(t)$. Moreover, the variables $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{v}}$ are reinitialized at the beginning of each time step with the values $\mathbf{q}_n$ and $\mathbf{v}_n$, i.e., $\tilde{\mathbf{q}}$ and $\tilde{\mathbf{v}}$ are discontinuous at t_n if an impact has occurred over the time step $(t_{n-1}, t_n]$. In this case, the algorithm, which relies on past values of $\dot{\tilde{\mathbf{v}}}(t)$, loses its accuracy properties and leads to $\mathcal{O}(h)$ errors in the velocity variable $\tilde{\mathbf{v}}(t)$. The proposed algorithm is nevertheless acceptable in the sense that $\mathcal{O}(h)$ errors are tolerated anyway for time steps with impacts and its numerical damping can efficiently stabilize the numerical but transient oscillations induced by these disturbances.
3. For numerical reasons, a prediction value $\mathbf{q}^*_{n+1}$ is used in Eq. (29h) instead of the actual value $\mathbf{q}_{n+1}$, similar to what is usually done in the Moreau–Jean scheme. In the following, the prediction $\mathbf{q}^*_{n+1} = \tilde{\mathbf{q}}_{n+1}$ shall be used. The reasons for this choice are discussed in Section 7.2.
4. Equations (29f) and (29h) for the velocity jump $\mathbf{W}_{n+1}$ and the multiplier $\mathbf{\Lambda}_{n+1}$ are similar to the conditions imposed in the classical Moreau–Jean method. The multiplier $\mathbf{\Lambda}_{n+1}$ is thus interpreted as the integral of the contact reaction forces over the time step $(t_n, t_{n+1}]$. Equation (29h) expresses the complementarity between the unilateral constraint at velocity level and the Lagrange multiplier Λ^j_{n+1}, which controls the amplitude of the velocity jump.

7. Properties and consistency of the scheme

7.1. Estimates of the Lagrange multipliers

We can derive an estimate of the multipliers $\boldsymbol{\nu}_{n+1}$ and $\mathbf{\Lambda}_{n+1}$. Let us define $\mathcal{A} \subset \mathcal{C}$ the set of constraints active at position level, which means that $\mathbf{g}^{\mathcal{A}}(\mathbf{q}_{n+1}) = \mathbf{0}$, and $\mathcal{B} \subset \mathcal{C}$ the set of constraints active at velocity level, which means that $\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n+1}\,\mathbf{v}_{n+1} + \mathbf{E}^{\mathcal{B}}\,\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n}\,\mathbf{v}_n = \mathbf{0}$, where $\mathbf{E}^{\mathcal{B}} = \mathrm{diag}\{e^j\}$ is the diagonal matrix containing the coefficients of restitution of the constraints in $\mathcal{B}$. In this expression, some coefficients of restitution have been artificially introduced for the bilateral constraints and are defined as $e^j = 0$, $\forall j \in \overline{\mathcal{U}}$. By construction, the bilateral constraints are in both $\mathcal{A}$ and $\mathcal{B}$.

Property 1. *If the mass matrix is nonsingular, we have*

$$\boldsymbol{\nu}^{\mathcal{A}}_{n+1} = -(\mathbf{g}^{\mathcal{A}}_{\mathbf{q},n+1}\mathbf{M}^{-1}_{n+1}\mathbf{g}^{\mathcal{A},T}_{\mathbf{q},n+1})^{-1}\,\mathbf{g}^{\mathcal{A}}(\tilde{\mathbf{q}}_{n+1}) + \mathcal{O}(h^2) \tag{32a}$$

$$\mathbf{\Lambda}^{\mathcal{B}}_{n+1} = -(\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n+1}\mathbf{M}^{-1}_{n+1}\mathbf{g}^{\mathcal{B},T}_{\mathbf{q},n+1})^{-1}\left(\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n+1}\tilde{\mathbf{v}}_{n+1} + \mathbf{E}^{\mathcal{B}}\,\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n}\mathbf{v}_n\right) \tag{32b}$$

Proof. We have $\mathbf{g}^T_{\mathbf{q},n+1}\ \boldsymbol{\Lambda}_{n+1} = \mathbf{g}^{\mathcal{B},T}_{\mathbf{q},n+1}\ \boldsymbol{\Lambda}^{\mathcal{B}}_{n+1}$ and $\mathbf{g}^T_{\mathbf{q},n+1}\ \boldsymbol{\nu}_{n+1} = \mathbf{g}^{\mathcal{A},T}_{\mathbf{q},n+1}\ \boldsymbol{\nu}^{\mathcal{A}}_{n+1}$. The proof follows (i) from the approximated form of the active constraints at position level $\mathbf{g}^{\mathcal{A}}(\mathbf{q}_{n+1}) = \mathbf{g}^{\mathcal{A}}(\tilde{\mathbf{q}}_{n+1}) + \mathbf{g}^{\mathcal{A}}_{\mathbf{q},n+1}\mathbf{U}_{n+1} + \mathcal{O}(h^2)$, which results from $\mathbf{U}_{n+1} = \mathcal{O}(h)$, with Eq. (29c) and (ii) from the exact form of the active constraints at velocity level $\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n+1}\,\mathbf{v}_{n+1} + \mathbf{E}^{\mathcal{B}}\,\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n}\,\mathbf{v}_n = \mathbf{g}^{\mathcal{B}}_{\mathbf{q},n+1}(\tilde{\mathbf{v}}_{n+1} + \mathbf{W}_{n+1}) + \mathbf{E}^{\mathcal{B}}\,\mathbf{g}^{\mathcal{B}}_{\mathbf{q},n}\mathbf{v}_n$ with Eq. (29f). □

Therefore, $\boldsymbol{\nu}_{n+1}$ can be seen as the set of multipliers that appears to compensate the violation of the constraints at position level by the free displacement; this quantity tends to converge to zero when the time step decreases and it is independent of the coefficient of restitution. Similarly, $\boldsymbol{\Lambda}_{n+1}$ can be seen as the set of multipliers that appears to compensate the violation of the constraints at velocity level, i.e., the Newton impact law, by the free velocity; this quantity depends on the value of the coefficient of restitution.

Based on Eq. (20a), the impulse of the constraint reaction forces can be estimated as

$$\int_{(t_n,t]} \mathrm{d}\mathbf{i} \simeq \boldsymbol{\Lambda}_{n+1} + h\tilde{\boldsymbol{\lambda}}_{n+1} \tag{33}$$

If impacts occur, $\boldsymbol{\Lambda}_{n+1} = \mathcal{O}(1)$ and the reaction forces are expected to be unbounded. Theorem 1 indicates that $\boldsymbol{\Lambda}_{n+1}$ is affected by an $\mathcal{O}(h)$ local error.

When all constraints remain either closed or open, i.e., no impact occurs, $\boldsymbol{\Lambda}_{n+1} = \mathcal{O}(h)$ and a more refined analysis reveals that this variable is only affected by an $\mathcal{O}(h^2)$ local error. The variable $(1/h)\boldsymbol{\Lambda}_{n+1} + \tilde{\boldsymbol{\lambda}}_{n+1}$ is thus a meaningful approximation the constraint reaction forces. Since $\tilde{\boldsymbol{\lambda}}^{\mathcal{U}} = \mathbf{0}$, the unilateral contact forces are represented by $(1/h)\boldsymbol{\Lambda}^{\mathcal{U}}_{n+1}$. Also, the generalized reaction forces of the bilateral constraints can be estimated as $(1/h)\boldsymbol{\Lambda}^{\overline{\mathcal{U}}}_{n+1} + \tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}}_{n+1}$. Illustrations of this estimate will be given in Section 10.2.

7.2. Choice of the prediction value for the activation of the velocity constraint

Different choices for the prediction $\mathbf{q}^*_{n+1}$ in Eq. (29h) could be considered such as $\mathbf{q}^*_{n+1} = \tilde{\mathbf{q}}_{n+1}$, $\mathbf{q}^*_{n+1} = \mathbf{q}_n$, $\mathbf{q}^*_{n+1} = \mathbf{q}_n + h\mathbf{v}_n$ or $\mathbf{q}^*_{n+1} = \mathbf{q}_{n+1}$, which would change the behaviour of the algorithm. It may seem attractive to use $\mathbf{q}^*_{n+1} = \mathbf{q}_{n+1}$ in order to prevent the activation of a constraint at velocity level when it is not active at position level. However, a tolerance on the criterion $g^j(\mathbf{q}^*_{n+1}) \leq tol$ is then needed where tol is defined as an upper bound on the numerical errors for the satisfaction of the constraint at position level. This tolerance would be particularly important for systems with nonlinear unilateral constraints, since any iterative Newton-type algorithm would only converge to the position level constraint up to certain tolerance. However, if the tolerance is well-chosen, the behavior of the algorithm would be similar as if $\mathbf{q}^*_{n+1} = \tilde{\mathbf{q}}_{n+1}$, because Eq. (32a) with the condition $\boldsymbol{\nu}^{\mathcal{A}}_{n+1} \geq \mathbf{0}$ imply that $\mathbf{g}^{\mathcal{A}}(\tilde{\mathbf{q}}_{n+1}) \leq \mathbf{0}$ for a sufficiently small time step. For that reason, the choice $\mathbf{q}^*_{n+1} = \tilde{\mathbf{q}}_{n+1}$ shall be used in the following since it does not require the selection of a tolerance.

7.3. Fundamental properties in a one degree-of-freedom test problem

In this section, the scheme is studied for the simple example of a bouncing ball, which can be seen as a single degree-of-freedom model equation for the contact problem.

The equations of motion of the bouncing ball take the form

$$\begin{aligned} \dot{q} &= v && \text{(34a)} \\ m\mathrm{d}v + \mathrm{d}i &= -ma_g\mathrm{d}t && \text{(34b)} \\ 0 \le q &\perp \mathrm{d}i \ge 0 && \text{(34c)} \end{aligned}$$

where m is the mass of the ball and a_g is the gravity acceleration.

According to the proposed time stepping scheme, the numerical solution is such that $\dot{\tilde{v}}_{n+1} = a_{n+1} = -a_g$ and

$$\begin{aligned} mU_{n+1} - \nu_{n+1} &= 0 && \text{(35a)} \\ 0 \le q_{n+1} \perp \nu_{n+1} &\ge 0 && \text{(35b)} \\ mW_{n+1} - \Lambda_{n+1} &= 0 && \text{(35c)} \\ \text{if } \tilde{q}_{n+1} \le 0 \text{ then } 0 \le v_{n+1} + e\, v_n \perp \Lambda_{n+1} &\ge 0 && \text{(35d)} \\ \tilde{q}_{n+1} &= q_n + hv_n - h^2 a_g/2 && \text{(35e)} \\ q_{n+1} &= \tilde{q}_{n+1} + U_{n+1} && \text{(35f)} \\ \tilde{v}_{n+1} &= v_n - ha_g && \text{(35g)} \\ v_{n+1} &= \tilde{v}_{n+1} + W_{n+1} && \text{(35h)} \end{aligned}$$

The next property is related with the equilibrium points of the numerical solution. Let us recall that a state q_n, v_n of the discrete system is an equilibrium point (or a fixed point) if we have $q_n = q_e$ and $v_n = 0\ \forall n$.

Property 2. *For the system represented by Eq.* (35)*, the state* $q_n = 0$*,* $v_n = 0$ *is the unique equilibrium point.*

Proof. Let us assume we are at an equilibrium point with $q_n = q_e$, $v_n = 0$. Then, we have

$$\begin{aligned} \tilde{q}_{n+1} &= q_e - h^2 a_g/2 \\ \tilde{v}_{n+1} &= -ha_g \end{aligned}$$

If $q_e > h^2 a_g/2$, the complementarity condition at velocity level is not activated so that $W_{n+1} = 0$, $v_{n+1} = -ha_g$ and the state is not an equilibrium point.

If $q_e \in [0, h^2 a_g/2)$, after some manipulations, the complementarity condition at velocity level becomes

$$0 \le -ha_g + W_{n+1} \perp mW_{n+1} \ge 0$$

which yields $W_{n+1} = ha_g$ so that $\Lambda_{n+1} = mha_g$ and $v_{n+1} = 0$, i.e., the constraint is active at velocity level. The complementarity condition at position level can be reformulated as

$$0 \leq q_e - h^2 a_g/2 + U_{n+1} \perp mU_{n+1} \geq 0$$

$\forall q_e \in [0, h^2 a_g/2)$, the solution is then obtained as

$$q_{n+1} = 0, \quad U_{n+1} \geq 0 \tag{36}$$

Any $q_e \neq 0$ yields $q_{n+1} = 0 \neq q_e$ and is thus not an equilibrium point. The value $q_e = 0$ yields $q_{n+1} = 0$ and is thus the unique equilibrium point. □

Property 3. *The numerical solution of the bouncing ball problem in Eq.* (34) *is such that the total energy decays monotonically*

$$E_{n+1} \leq E_n$$

where $E_n = mv_n^2/2 + ma_g q_n$.

Proof. The proof relies on the analysis of the different activation scenarios of the unilateral constraint.

The first situation is the free flight phase. Since the smooth motion is solved exactly, the energy is exactly preserved by the algorithm.

The second situation is a simultaneous activation of the constraint at position and velocity levels, which implies $v_{n+1} = -e\, v_n$ and $q_{n+1} = 0$. The variation of energy is then equal to

$$E_{n+1} - E_n = -q_n m a_g - \frac{1}{2} m(1 - e^2) v_n^2$$

Using $q_n \geq 0$ and $e \in [0, 1]$, we obtain $E_{n+1} - E_n \leq 0$.

The third situation is such that the constraint is not active at velocity level because $\tilde{v}_{n+1} + e\, v_n > 0$ but is active at position level because $\tilde{q}_{n+1} < 0$. We would have simultaneously

$$\begin{aligned} \tilde{q}_{n+1} = q_n + h\, v_n - \frac{1}{2} h^2\, a_g &< 0 \\ \tilde{v}_{n+1} + e\, v_n = v_n - h\, a_g + e\, v_n &> 0 \end{aligned}$$

which would imply that

$$\frac{h^2\, a_g}{1 + e} < h\, v_n < \frac{h^2\, a_g}{2} - q_n$$

Since $e \in [0, 1]$ and $q_n \geq 0$, these conditions can never be satisfied so that this situation cannot be encountered.

The fourth situation is such that the constraint is active at velocity level and not at position level. This situation is forbidden in the proposed algorithm. □

8. Newton semi-smooth method

This section addresses the solution of the discrete nonlinear system (29,30) for a given time step $n+1$, using a Newton procedure with an active set strategy. The algorithm is described in the special case $\mathbf{q}^*_{n+1} = \tilde{\mathbf{q}}_{n+1}$.

The complementarity conditions in Eq. (29) can be reformulated to get a nonlinear semi-smooth problem as follows

$$
\begin{aligned}
\mathbf{M}(\mathbf{q}_{n+1})\dot{\tilde{\mathbf{v}}}_{n+1} - \mathbf{f}(\mathbf{q}_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) - \mathbf{g}^{\overline{\mathcal{U}},T}_{\mathbf{q},n+1}\,\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}}_{n+1} &= \mathbf{0} \quad (37a)\\
\mathbf{g}^{\overline{\mathcal{U}}}_{\mathbf{q},n+1}\,\tilde{\mathbf{v}}_{n+1} &= \mathbf{0} \quad (37b)\\
\mathbf{M}(\mathbf{q}_{n+1})\mathbf{U}_{n+1} - \mathbf{g}^{T}_{\mathbf{q},n+1}\,\boldsymbol{\nu}_{n+1} &= \mathbf{0} \quad (37c)\\
\mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}_{n+1}) &= \mathbf{0} \quad (37d)\\
\boldsymbol{\nu}^{\mathcal{U}}_{n+1} - \max(\mathbf{0}, \boldsymbol{\nu}^{\mathcal{U}}_{n+1} - r\mathbf{g}^{\mathcal{U}}(\mathbf{q}_{n+1})) &= \mathbf{0} \quad (37e)\\
\mathbf{M}(\mathbf{q}_{n+1})\mathbf{W}_{n+1} - \mathbf{g}^{T}_{\mathbf{q},n+1}\boldsymbol{\Lambda}_{n+1} &= \mathbf{0} \quad (37f)\\
\mathbf{g}^{\overline{\mathcal{U}}}_{\mathbf{q},n+1}\mathbf{v}_{n+1} &= \mathbf{0} \quad (37g)\\
\text{if } g^j(\tilde{\mathbf{q}}_{n+1}) \le 0 \text{ then } \Lambda^j - \max(0, \Lambda^j - r(g^j_{\mathbf{q},n+1}\,\mathbf{v}_{n+1} + e^j\, g^j_{\mathbf{q},n}\,\mathbf{v}_n)) &= 0, \quad \forall j \in \mathcal{U} \quad (37h)
\end{aligned}
$$

with the numerical coefficient $r > 0$. This nonlinear semi-smooth problem leads to a monolithic scheme to iteratively compute the solution and contact conditions. The solution of this semi-smooth problem is used to advance the computation at each time step.

The active sets $\mathcal{A}_{n+1}$ and $\mathcal{B}_{n+1}$ and their complementarity sets $\overline{\mathcal{A}}_{n+1}$ and $\overline{\mathcal{B}}_{n+1}$ are defined as the sets of constraints

$$
\begin{aligned}
\mathcal{A}_{n+1} &= \overline{\mathcal{U}} \cup \left\{ j \in \mathcal{U} : \nu^j_{n+1} - r\, g^j(\mathbf{q}_{n+1}) \ge 0 \right\} && (38a)\\
\overline{\mathcal{A}}_{n+1} &= \mathcal{C} \setminus \mathcal{A}_{n+1} && (38b)\\
\mathcal{B}_{n+1} &= \overline{\mathcal{U}} \cup \big\{ j \in \mathcal{U} : g^j(\tilde{\mathbf{q}}_{n+1}) \le 0 \text{ and} \\
&\qquad \Lambda^j_{n+1} - r(g^j_{\mathbf{q},n+1}\mathbf{v}_{n+1} + e^j\, g^j_{\mathbf{q},n}\mathbf{v}_n) \ge 0) \big\} && (38c)\\
\overline{\mathcal{B}}_{n+1} &= \mathcal{C} \setminus \mathcal{B}_{n+1} && (38d)
\end{aligned}
$$

One observes that these two activation criteria are completely implicit and have a similar structure as the activation criterion proposed by Alart and Curnier [24]

for quasi-static problems. The scaled vectors of residuals are defined as follows:

$$\mathbf{r}^s = \begin{bmatrix} \mathbf{M}(\mathbf{q}_{n+1})h\dot{\tilde{\mathbf{v}}}_{n+1} - \mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}},T}\, h\tilde{\boldsymbol{\lambda}}_{n+1}^{\overline{\mathcal{U}}} - h\mathbf{f}(\mathbf{q}_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) \\ \mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}}}\, \tilde{\mathbf{v}}_{n+1} \end{bmatrix} \tag{39a}$$

$$\mathbf{r}^p = \frac{1}{h}\begin{bmatrix} \mathbf{M}(\mathbf{q}_{n+1})\mathbf{U}_{n+1} - \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{A},T}\, \boldsymbol{\nu}_{n+1}^{\mathcal{A}} \\ \mathbf{g}^{\mathcal{A}}(\mathbf{q}_{n+1}) \\ \boldsymbol{\nu}_{n+1}^{\overline{\mathcal{A}}} \end{bmatrix} \tag{39b}$$

$$\mathbf{r}^v = \begin{bmatrix} \mathbf{M}(\mathbf{q}_{n+1})\mathbf{W}_{n+1} - \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{B},T}\boldsymbol{\Lambda}_{n+1}^{\mathcal{B}} \\ \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{B}}\, \mathbf{v}_{n+1} + \mathbf{E}^{\mathcal{B}}\, \mathbf{g}_{\mathbf{q},n}^{\mathcal{B}}\, \mathbf{v}_n \\ \boldsymbol{\Lambda}^{\overline{\mathcal{B}}} \end{bmatrix} \tag{39c}$$

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}^s \\ \mathbf{r}^p \\ \mathbf{r}^v \end{bmatrix} \tag{39d}$$

In the above expressions, $\mathcal{A}_{n+1}$ and $\mathcal{B}_{n+1}$ are denoted by $\mathcal{A}$ and $\mathcal{B}$ for the sake of readability, whereas the matrix of restitution coefficients $\mathbf{E}^{\mathcal{B}}$ was introduced in Section 7.1. It can be observed that the numerical coefficient r only appears in the activation criteria but not in the residuals (39), which is an important difference compared to the Augmented Lagrangian formulation described in [24]. Let us remark that the residuals could be reorganized in a form similar to an Augmented Lagrangian formulation by the addition of penalty terms, i.e., by replacing the terms

$$\mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}},T}\, h\tilde{\boldsymbol{\lambda}}_{n+1}^{\overline{\mathcal{U}}},\ \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{A},T}\, \boldsymbol{\nu}_{n+1}^{\mathcal{A}},\ \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{B},T}\boldsymbol{\Lambda}_{n+1}^{\mathcal{B}}$$

by

$$\begin{aligned} &\mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}},T}\, h(\tilde{\boldsymbol{\lambda}}_{n+1}^{\overline{\mathcal{U}}} - r\, \mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}}}\, \tilde{\mathbf{v}}_{n+1}),\ \mathbf{g}_{\mathbf{q},n+1}^{\mathcal{A},T}\, (\boldsymbol{\nu}_{n+1}^{\mathcal{A}} - r\mathbf{g}^{\mathcal{A}}(\mathbf{q}_{n+1})), \\ &\mathbf{g}_{\mathbf{q},n+1}^{\mathcal{B},T}(\boldsymbol{\Lambda}_{n+1}^{\mathcal{B}} - r(\mathbf{g}_{\mathbf{q},n+1}^{\mathcal{B}}\, \mathbf{v}_{n+1} + \mathbf{E}^{\mathcal{B}}\, \mathbf{g}_{\mathbf{q},n}^{\mathcal{B}}\, \mathbf{v}_n)), \end{aligned}$$

respectively. These penalty terms would not modify the numerical solution but they could tend to bring some convexity to the problem, thereby improving the convergence of the Newton semi-smooth iterations. However, this reformulation was not necessary in the test cases described below.

The prediction is based on the initial guesses $\dot{\tilde{\mathbf{v}}}_{n+1}^0 = \mathbf{0}$, $\mathbf{W}_{n+1}^0 = \mathbf{0}$, $\mathbf{U}_{n+1}^0 = \mathbf{0}$,

and $\dot{\mathbf{q}}_n = \mathbf{v}_n$, which implies:

$$\begin{aligned}
\mathbf{q}^0_{n+1} &= \mathbf{q}_n + h\mathbf{v}_n + h^2(0.5-\beta)\mathbf{a}_n + h^2\beta(\alpha_f\dot{\tilde{\mathbf{v}}}_n - \alpha_m\mathbf{a}_n)/(1-\alpha_m) && \text{(40a)}\\
\tilde{\mathbf{v}}^0_{n+1} &= \mathbf{v}_n + h(1-\gamma)\mathbf{a}_n + h\gamma(\alpha_f\dot{\tilde{\mathbf{v}}}_n - \alpha_m\mathbf{a}_n)/(1-\alpha_m) && \text{(40b)}\\
\mathbf{a}^0_{n+1} &= (\alpha_f\dot{\tilde{\mathbf{v}}}_n - \alpha_m\mathbf{a}_n)/(1-\alpha_m) && \text{(40c)}\\
\mathbf{v}^0_{n+1} &= \tilde{\mathbf{v}}^0_{n+1} && \text{(40d)}\\
\mathbf{U}^0_{n+1} &= \mathbf{0} && \text{(40e)}\\
\mathbf{W}^0_{n+1} &= \mathbf{0} && \text{(40f)}\\
\tilde{\boldsymbol{\lambda}}^0_{n+1} &= \mathbf{0} && \text{(40g)}\\
\boldsymbol{\nu}^0_{n+1} &= \mathbf{0} && \text{(40h)}\\
\boldsymbol{\Lambda}^0_{n+1} &= \mathbf{0} && \text{(40i)}
\end{aligned}$$

Then, the correction terms should satisfy the difference formulae, so that the corrections $\Delta\dot{\tilde{\mathbf{v}}}_{n+1}$ and $\Delta\mathbf{q}_n$ can be eliminated in terms of the corrections $\Delta\tilde{\mathbf{v}}_{n+1}$, $\Delta\mathbf{W}_{n+1}$ and $\Delta\mathbf{U}_{n+1}$.

$$\begin{aligned}
\Delta\mathbf{v}_{n+1} &= \Delta\tilde{\mathbf{v}}_{n+1} + \Delta\mathbf{W}_{n+1} && \text{(41a)}\\
\Delta\dot{\tilde{\mathbf{v}}}_{n+1} &= (1-\alpha_m)/((1-\alpha_f)\gamma h)\,\Delta\tilde{\mathbf{v}}_{n+1} && \text{(41b)}\\
\Delta\mathbf{q}_{n+1} &= h\,\beta/\gamma\,\Delta\tilde{\mathbf{v}}_{n+1} + \Delta\mathbf{U}_{n+1} && \text{(41c)}
\end{aligned}$$

Introducing the vector of independent corrections

$$\Delta\mathbf{x} = \begin{bmatrix} \Delta\tilde{\mathbf{v}} \\ -h\Delta\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}} \\ \Delta\mathbf{U}/h \\ -\Delta\boldsymbol{\nu}^{\mathcal{A}}/h \\ -\Delta\boldsymbol{\nu}^{\overline{\mathcal{A}}}/h \\ \Delta\mathbf{W} \\ -\Delta\boldsymbol{\Lambda}^{\mathcal{B}} \\ -\Delta\boldsymbol{\Lambda}^{\overline{\mathcal{B}}} \end{bmatrix} \tag{42}$$

the correction equation is obtained as

$$\mathbf{S}_t\,\Delta\mathbf{x} = -\mathbf{r} \tag{43}$$

In this expression, the iteration matrix is defined as

$$\mathbf{S}_t = \left[\begin{array}{cc|ccc|ccc}
\mathbf{S}^*_t & \mathbf{g}^{\overline{\mathcal{U}},T}_{\mathbf{q}} & h^2\mathbf{K}_t & \mathbf{0} & \mathbf{0} & h\mathbf{C}_t & \mathbf{0} & \mathbf{0} \\
\mathbf{g}^{\overline{\mathcal{U}}}_{\mathbf{q}} + \frac{h\beta}{\gamma}\mathbf{G}^{*s} & \mathbf{0} & h\mathbf{G}^{*s} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
\frac{\beta}{\gamma}\mathbf{G}^p & \mathbf{0} & \mathbf{M}+\mathbf{G}^p & \mathbf{g}^{\mathcal{A},T}_{\mathbf{q}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\frac{\beta}{\gamma}\mathbf{g}^{\mathcal{A}}_{\mathbf{q}} & \mathbf{0} & \mathbf{g}^{\mathcal{A}}_{\mathbf{q}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{A}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\hline
\frac{h\beta}{\gamma}\mathbf{G}^v & \mathbf{0} & h\mathbf{G}^v & \mathbf{0} & \mathbf{0} & \mathbf{M} & \mathbf{g}^{\mathcal{B},T}_{\mathbf{q}} & \mathbf{0} \\
\mathbf{g}^{\mathcal{B}}_{\mathbf{q}} + \frac{h\beta}{\gamma}\mathbf{G}^{*v} & \mathbf{0} & h\mathbf{G}^{*v} & \mathbf{0} & \mathbf{0} & \mathbf{g}^{\mathcal{B}}_{\mathbf{q}} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{B}}}
\end{array}\right] \tag{44}$$

where $\mathbf{I}^{\overline{\mathcal{A}}}$ and $\mathbf{I}^{\overline{\mathcal{B}}}$ are identity matrices and

$$\begin{aligned}
&\mathbf{S}_t^* = \frac{1-\alpha_m}{(1-\alpha_f)\gamma}\mathbf{M} + h\mathbf{C}_t + \frac{\beta h^2}{\gamma}\mathbf{K}_t, \quad \mathbf{K}_t = \frac{\partial(\mathbf{M}(\mathbf{q})\dot{\tilde{\mathbf{v}}} - \mathbf{f}(\mathbf{q},\mathbf{v},t) - \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T}\tilde{\boldsymbol{\lambda}}^{\overline{\mathcal{U}}})}{\partial \mathbf{q}},\\
&\mathbf{C}_t = -\frac{\partial \mathbf{f}(\mathbf{q},\mathbf{v},t)}{\partial \mathbf{v}}, \quad \mathbf{G}^p = \frac{\partial(\mathbf{M}(\mathbf{q})\mathbf{U} - \mathbf{g}_{\mathbf{q}}^{\mathcal{A},T}\boldsymbol{\nu}^{\mathcal{A}})}{\partial \mathbf{q}}, \quad \mathbf{G}^v = \frac{\partial(\mathbf{M}(\mathbf{q})\mathbf{W} - \mathbf{g}_{\mathbf{q}}^{\mathcal{B},T}\boldsymbol{\Lambda}^{\mathcal{B}})}{\partial \mathbf{q}},\\
&\mathbf{G}^{*s} = \frac{\partial(\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}}\tilde{\mathbf{v}})}{\partial \mathbf{q}}, \quad \mathbf{G}^{*v} = \frac{\partial(\mathbf{g}_{\mathbf{q}}^{\mathcal{B}}\mathbf{v})}{\partial \mathbf{q}}
\end{aligned} \tag{45}$$

with the tangent stiffness matrix $\mathbf{K}_t$ and the tangent damping matrix $\mathbf{C}_t$. The matrices $\mathbf{G}^p$, $\mathbf{G}^v$, $\mathbf{G}^{*s}$ and $\mathbf{G}^{*v}$ are related with the second derivatives of the constraints and they vanish for linear constraints. The scaling of the multipliers in Eq. (42) by negative numbers has been selected in order to obtain symmetric contributions in the iteration matrix $\mathbf{S}_t$ and to ensure that this matrix does not become ill-conditioned for small time steps, see, e.g., [25] for a more detailed discussion on DAE scaling strategies. Indeed, since $\mathbf{U} = \mathcal{O}(h)$ and $\boldsymbol{\nu} = \mathcal{O}(h)$, we have $\mathbf{G}^p = \mathcal{O}(h)$ and one can observe that

$$\lim_{h\to 0}\mathbf{S}_t = \left[\begin{array}{cc|ccccc|ccc}
\frac{1-\alpha_m}{(1-\alpha_f)\gamma}\mathbf{M} & \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\hline
\frac{\beta}{\gamma}\mathbf{G}^p & \mathbf{0} & \mathbf{M} & \mathbf{g}_{\mathbf{q}}^{\mathcal{A},T} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\frac{\beta}{\gamma}\mathbf{g}_{\mathbf{q}}^{\mathcal{A}} & \mathbf{0} & \mathbf{g}_{\mathbf{q}}^{\mathcal{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{A}}} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
\hline
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M} & \mathbf{g}_{\mathbf{q}}^{\mathcal{B},T} & \mathbf{0}\\
\mathbf{g}_{\mathbf{q}}^{\mathcal{B}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{g}_{\mathbf{q}}^{\mathcal{B}} & \mathbf{0} & \mathbf{0}\\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{B}}}
\end{array}\right] \tag{46}$$

is a non-singular and well-conditioned matrix if we assume that the matrix in Eq (9) is also well-conditioned.

If the tangent damping and stiffness matrices are equal to $\mathbf{0}$ (which means that the smooth force vector $\mathbf{f}$ does not depend on $\mathbf{q}$ or $\mathbf{v}$) and if the bilateral constraints are linear (which implies $\mathbf{G}^{*s} = \mathbf{0}$), a lower-triangular block matrix is obtained and the system can be solved based on a sequential solution of the three subproblems. In the general case, this sequential approach also converges in many practical situations, even if the coupling terms are neglected at each iteration. This sequential approach can simplify the implementation in a general purpose code and improve the efficiency of the simulation. The complete procedure is

described in Algorithm 1 and involves the following matrices

$$\mathbf{S}_t^s = \begin{bmatrix} \mathbf{S}_t^* & \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}},T} \\ \mathbf{g}_{\mathbf{q}}^{\overline{\mathcal{U}}} + \frac{h\beta}{\gamma}\mathbf{G}^{*s} & \mathbf{0} \end{bmatrix} \tag{47}$$

$$\mathbf{S}_t^p = \begin{bmatrix} \mathbf{M} + \mathbf{G}^p & \mathbf{g}_{\mathbf{q}}^{\mathcal{A},T} & \mathbf{0} \\ \mathbf{g}_{\mathbf{q}}^{\mathcal{A}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{A}}} \end{bmatrix} \tag{48}$$

$$\mathbf{S}_t^v = \begin{bmatrix} \mathbf{M} & \mathbf{g}_{\mathbf{q}}^{\mathcal{B},T} & \mathbf{0} \\ \mathbf{g}_{\mathbf{q}}^{\mathcal{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}^{\overline{\mathcal{B}}} \end{bmatrix} \tag{49}$$

It can be summarized as follows. At each iteration of the Newton procedure, Algorithm 1 involves 3 steps:

1. solve the linearized equation of the smooth motion (i.e., all contributions of unilateral constraints are disregarded) using the classical procedure obtained from the generalized-α scheme for smooth systems with the bilateral constraints expressed at velocity level;
2. project on the position constraints in $\mathcal{A}$ using the matrix $\mathbf{S}_t^p$;
3. project on the velocity constraints in $\mathcal{B}$ using the matrix $\mathbf{S}_t^v$.

If convergence difficulties are encountered, one may consider either to reduce the time step, which tends to reduce the influence of the coupling terms $h^2\mathbf{K}_t$, $h\mathbf{G}^{*s}$, and $h\mathbf{C}_t$, or to solve the coupled problem based on the full matrix $\mathbf{S}_t$ in Eq. (44).

9. Alternative schemes

In the numerical examples below, the proposed nonsmooth generalized-α GGL method is compared with other two algorithms from the literature, which are briefly described in this section.

9.1. Moreau–Jean θ-method

Assuming that the mass matrix is non-singular, the Moreau–Jean θ-method can be formulated based on the following discrete set of equations

$$\mathbf{M}(\mathbf{q}_{n+1})\dot{\tilde{\mathbf{v}}}_{n+1} - \mathbf{f}(\mathbf{q}_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) = \mathbf{0} \tag{50a}$$

$$\mathbf{M}(\mathbf{q}_{n+1})\mathbf{W}_{n+1} - \mathbf{g}_{\mathbf{q},n+1}^T \mathbf{\Lambda}_{n+1} = \mathbf{0} \tag{50b}$$

$$\mathbf{g}_{\mathbf{q},n+1}^{\overline{\mathcal{U}}} \mathbf{v}_{n+1} = \mathbf{0} \tag{50c}$$

$$\text{if } g^j(\mathbf{q}_{n+1}^*) \le 0 \text{ then } \Lambda^j - \max(0, \Lambda^j - r(g_{\mathbf{q},n+1}^j \mathbf{v}_{n+1} + e^j\, g_{\mathbf{q},n}^j \mathbf{v}_n)) = 0, \quad \forall j \in \mathcal{U} \tag{50d}$$

with the integration formulae

$$\tilde{\mathbf{q}}_{n+1} = \mathbf{q}_n + h(1-\theta)\mathbf{v}_n + h\theta\tilde{\mathbf{v}}_{n+1} \tag{50e}$$

$$\tilde{\mathbf{v}}_{n+1} = \mathbf{v}_n + h(1-\theta)\dot{\tilde{\mathbf{v}}}_n + h\theta\dot{\tilde{\mathbf{v}}}_{n+1} \tag{50f}$$

$$\mathbf{v}_{n+1} = \tilde{\mathbf{v}}_{n+1} + \mathbf{W}_{n+1} \tag{50g}$$

$$\mathbf{q}_{n+1} = \tilde{\mathbf{q}}_{n+1} + h\theta\mathbf{W}_{n+1} \tag{50h}$$

Algorithm 1 Nonsmooth GGL generalized-α time integration scheme

Inputs: initial values $\mathbf{q}_0$ and $\mathbf{v}_0$
Compute the consistent value of $\dot{\tilde{\mathbf{v}}}_0$
$\mathbf{a}_0 := \dot{\tilde{\mathbf{v}}}_0$
for $n = 0$ to $n_{\text{final}} - 1$ **do**
  $\dot{\tilde{\mathbf{v}}}_{n+1} := \mathbf{0}$, $\tilde{\boldsymbol{\lambda}}^{\overline{U}}_{n+1} := \mathbf{0}$, $\boldsymbol{\nu}_{n+1} := \mathbf{0}$, $\boldsymbol{\Lambda}_{n+1} := \mathbf{0}$
  $\mathbf{a}_{n+1} := 1/(1-\alpha_m)(\alpha_f \dot{\tilde{\mathbf{v}}}_n - \alpha_m \mathbf{a}_n)$
  $\widetilde{\mathbf{v}}_{n+1} := \mathbf{v}_n + h(1-\gamma)\mathbf{a}_n + h\gamma \mathbf{a}_{n+1}$
  $\mathbf{v}_{n+1} := \widetilde{\mathbf{v}}_{n+1}$
  $\mathbf{q}_{n+1} := \mathbf{q}_n + h\mathbf{v}_n + h^2(1/2\ - \beta)\mathbf{a}_n + h^2\beta \mathbf{a}_{n+1}$
  for $i = 1$ to i_{max} **do**
    Compute $\mathbf{r}^s$, $\mathbf{r}^p$ and $\mathbf{r}^v$ using Eqs. (39a-39c)
    if $\|\mathbf{r}^s\| < tol^s$ and $\|\mathbf{r}^p\| < tol^p$ and $\|\mathbf{r}^v\| < tol^v$ **then**
      break
    end if
    Step 1 (smooth motion):
      Compute the iteration matrix $\mathbf{S}^s_t$ using Eq. (47)
      $\begin{bmatrix} \Delta\tilde{\mathbf{v}} \\ -h\Delta\tilde{\boldsymbol{\lambda}}^{\overline{U}} \end{bmatrix} := -(\mathbf{S}^s_t)^{-1}\mathbf{r}^s$
      $\tilde{\mathbf{v}}_{n+1} := \tilde{\mathbf{v}}_{n+1} + \Delta\tilde{\mathbf{v}}$
      $\dot{\tilde{\mathbf{v}}}_{n+1} := \dot{\tilde{\mathbf{v}}}_{n+1} + (1-\alpha_m)/((1-\alpha_f)\gamma h)\Delta\tilde{\mathbf{v}}$
      $\mathbf{v}_{n+1} := \tilde{\mathbf{v}}_{n+1} + \mathbf{W}_{n+1}$
      $\mathbf{q}_{n+1} := \mathbf{q}_{n+1} + h\beta/\gamma\Delta\tilde{\mathbf{v}}$
      $\widetilde{\boldsymbol{\lambda}}^{\overline{U}}_{n+1} := \widetilde{\boldsymbol{\lambda}}^{\overline{U}}_{n+1} + \Delta\widetilde{\boldsymbol{\lambda}}^{\overline{U}}$
    Step 2 (projection on position constraints):
      Compute $\mathbf{r}^p$ using Eq. (39b)
      Compute $\mathbf{S}^p_t$ using Eq. (48)
      $\begin{bmatrix} \Delta\mathbf{U} \\ -\Delta\boldsymbol{\nu}/h \end{bmatrix} := -(\mathbf{S}^p_t)^{-1}\mathbf{r}^p$
      $\mathbf{U}_{n+1} := \mathbf{U}_{n+1} + \Delta\mathbf{U}$
      $\mathbf{q}_{n+1} := \mathbf{q}_{n+1} + \Delta\mathbf{U}$
      $\boldsymbol{\nu}_{n+1} := \boldsymbol{\nu}_{n+1} + \Delta\boldsymbol{\nu}$
    Step 3 (projection on velocity constraints):
      Compute $\mathbf{r}^v$ using Eq. (39c)
      Compute $\mathbf{S}^v_t$ using Eq. (49)
      $\begin{bmatrix} \Delta\mathbf{W} \\ -\Delta\boldsymbol{\Lambda} \end{bmatrix} := -(\mathbf{S}^v_t)^{-1}\mathbf{r}^v$
      $\mathbf{W}_{n+1} := \mathbf{W}_{n+1} + \Delta\mathbf{W}$
      $\mathbf{v}_{n+1} := \tilde{\mathbf{v}}_{n+1} + \mathbf{W}_{n+1}$
      $\boldsymbol{\Lambda}_{n+1} := \boldsymbol{\Lambda}_{n+1} + \Delta\boldsymbol{\Lambda}$
  end for
  $\mathbf{a}_{n+1} := \mathbf{a}_{n+1} + (1-\alpha_f)/(1-\alpha_m)\dot{\tilde{\mathbf{v}}}_{n+1}$
end for

This system of equations can be solved using a Newton method, solving the complementarity condition in Eq. (50d) at every iteration. It is important that the activation of the complementarity condition in Eq. (50d) relies on an evaluation of the constraint g^j at the position $\mathbf{q}^*_{n+1}$, where $\mathbf{q}^*_{n+1}$ is an explicit prediction of the position which is not updated during the Newton iteration process. The reason is that, if the activation was based on the constraint at position level at time t_{n+1}, i.e., on the sign of $g^j(\mathbf{q}_{n+1})$, the discrete problem may not always be solvable and the Newton iterations may fail to converge. As a consequence, the Moreau–Jean scheme does not provide any guarantee on the sign of the unilateral constraint at position-level for the final response.

9.2. Smooth generalized-α index-3 method

This algorithm relies on a formulation of the constraints at position level only. It assumes that there is no jump in the velocities and is therefore only consistent if no impact occurs.

$$\begin{aligned} \mathbf{M}(\mathbf{q}_{n+1})\dot{\mathbf{v}}_{n+1} - \mathbf{g}^T_{\mathbf{q},n+1}\boldsymbol{\lambda}_{n+1} &= \mathbf{f}(\mathbf{q}_{n+1}, \mathbf{v}_{n+1}, t_{n+1}) && \text{(51a)} \\ \mathbf{g}^{\overline{\mathcal{U}}}(\mathbf{q}_{n+1}) &= \mathbf{0} && \text{(51b)} \\ \boldsymbol{\lambda}^{\mathcal{U}}_{n+1} - \max(\mathbf{0}, \boldsymbol{\lambda}^{\mathcal{U}}_{n+1} - r\mathbf{g}^{\mathcal{U}}(\mathbf{q}_{n+1})) &= \mathbf{0} && \text{(51c)} \end{aligned}$$

which are combined with the integration formulae

$$\begin{aligned} \mathbf{q}_{n+1} &= \mathbf{q}_n + h\mathbf{v}_n + h^2(0.5-\beta)\mathbf{a}_n + h^2\beta\mathbf{a}_{n+1} && \text{(52a)} \\ \tilde{\mathbf{v}}_{n+1} &= \mathbf{v}_n + h(1-\gamma)\mathbf{a}_n + h\gamma\mathbf{a}_{n+1} && \text{(52b)} \\ (1-\alpha_m)\mathbf{a}_{n+1} + \alpha_m\mathbf{a}_n &= (1-\alpha_f)\dot{\mathbf{v}}_{n+1} + \alpha_f\dot{\mathbf{v}}_n && \text{(52c)} \end{aligned}$$

It is noticeable that this algorithm does not make use of the impact law, i.e., the value of the coefficient of restitution is not needed to obtain a numerical solution.

10. Numerical examples

In order to highlight the properties of the proposed time integration method, several examples are studied as shown in Figure 1. The bouncing ball and the elastic bar have a linear behaviour in free motion and can be compared to exact solutions. The impact of a simple rigid pendulum illustrates the behaviour of the scheme when unilateral and bilateral constraints are simultaneously present. The rocking block example is modelled using two nonlinear unilateral constraints and its motion exhibits multiple impacts.

Several algorithms are compared:

- Smooth-α index-3: the generalized-α scheme with constraints at position level only, i.e., no coefficient of restitution is used, the numerical coefficients are defined according to [22] in order to have a spectral radius $\rho_\infty = 0.8$,

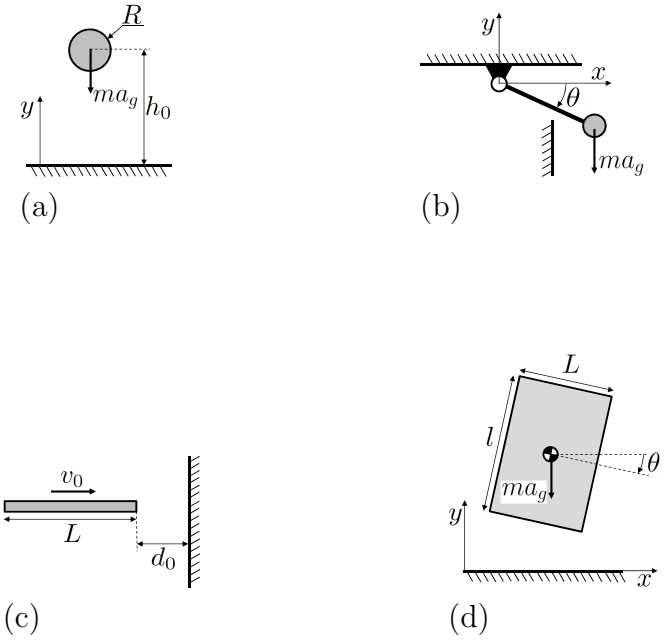

Figure 1: Examples: (a) bouncing ball; (b) rigid pendulum; (c) elastic bar; (d) rocking block.

- Nonsmooth-α, modified LCP: a variant of the proposed scheme when the assumption represented by Eq. (28) is relaxed and the complementarity condition (29e) is replaced by Eq. (25) with the approximation of $\boldsymbol{\mu}^{\mathcal{U}}$ in Eq. (27), the numerical coefficients are defined according to [22] in order to have a spectral radius $\rho_\infty = 0.8$,

- Moreau: the Moreau θ-method, where θ is selected in order to have a spectral radius $\rho_\infty = 0.8$ using the formula $\theta = 1/(\rho_\infty + 1)$,

- Nonsmooth-α GGL: the proposed GGL approach with constraints at velocity and position level, the numerical coefficients are defined according to [22] in order to have a spectral radius $\rho_\infty = 0.8$,

- Reference: a reference solution computed either with Nonsmooth-α GGL scheme and a smaller time step, or analytically.

10.1. The bouncing ball

The first example is a standard bouncing rigid ball on a rigid plane, as depicted in Figure 1(a). The ball starts from rest, falls due to the gravity and rebounds against the rigid plane with a restitution coefficient $e = 0.8$. Thus, the model includes a unilateral constraint on the vertical position: $0 \leq g^{\mathcal{U}} = y - R \perp \mathrm{d}i^{\mathcal{U}} \geq 0$, where R is the radius of the ball. The dynamic response displays an accumulation phenomena with an infinite number of rebounds over a finite time interval. An event-driven technique would fail to simulate this phenomenon since it is not possible to stop and restart the numerical integration an infinite

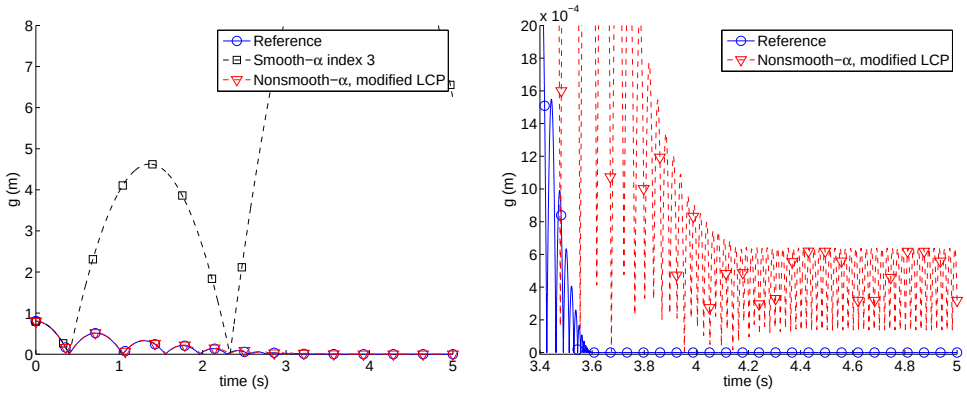

Figure 2: Bouncing ball: gap distance and zoom.

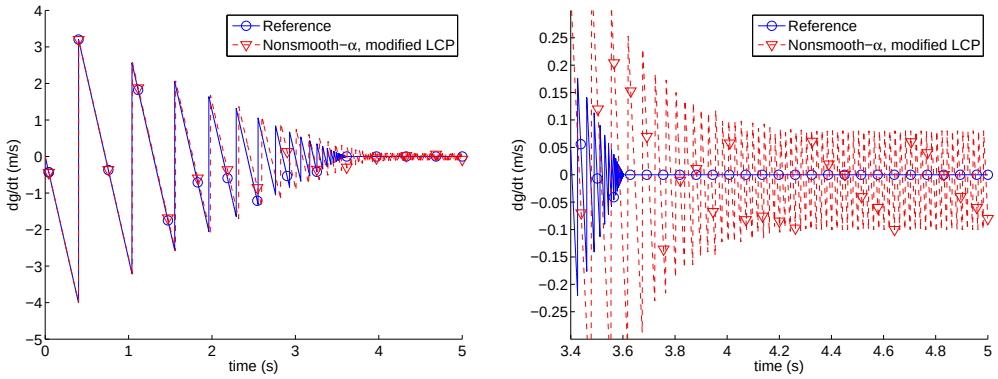

Figure 3: Bouncing ball: velocity and zoom.

number of times. The physical parameters of this model are as follows: mass $m = 1$ kg, radius $R = 0.2$ m, gravity acceleration $a_g = 10$ m/s^2, initial height $h_0 = 1.001$ m. With an initial height of $h_0 = 1$ m, the first impact would be synchronized with a time step; this special case would be numerically favorable but not representative of a general situation; this is the reason why the value $h_0 = 1.001$ m is selected. The numerical solution is computed with a time step $h = 2.\,10^{-3}$ s. The reference solution has been obtained analytically and the accumulation instant is $t = 3.602$ s.

In Figure 2, one observes that the smooth algorithm leads to an erroneous dynamic response characterized by increases in the energy at the impact times. Therefore, this algorithm is not consistent in the presence of impacts and it will not be studied in more detail for this example. The variant of the Nonsmooth-α scheme with the modified LCP condition based on Eq. (27) is then tested. The numerical solution behaves well until the accumulation phenomenon. Then, the constraints at position level (Figure 2) and at velocity level (Figure 3) oscillate between active and inactive states. A similar chattering behaviour was already investigated in [12]. This observation demonstrates that, when the assumption represented by Eq. (28) is not used, inconsistent results can be obtained.

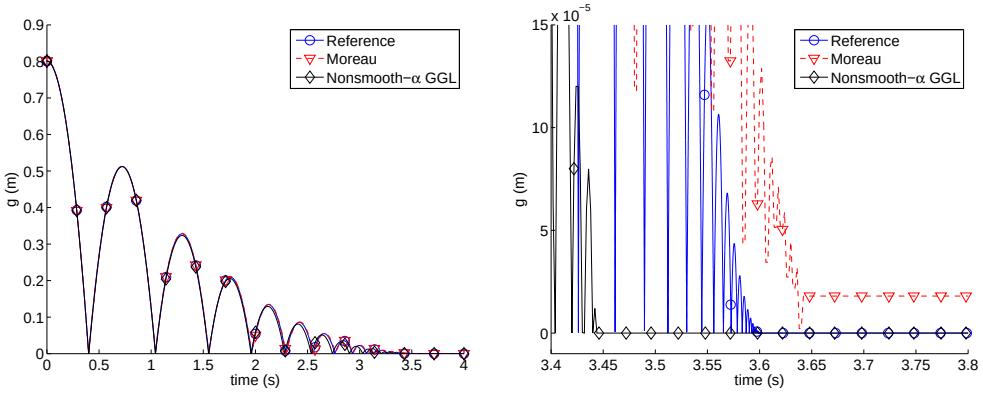

Figure 4: Bouncing ball: gap distance and zoom.

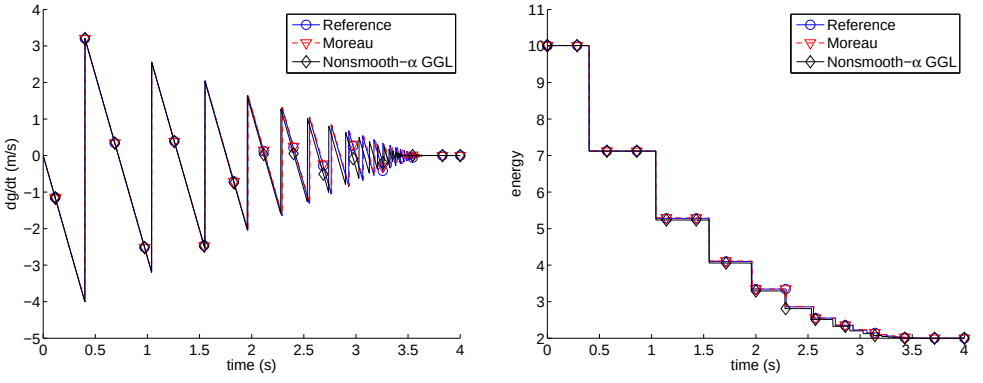

Figure 5: Bouncing ball: velocity and energy.

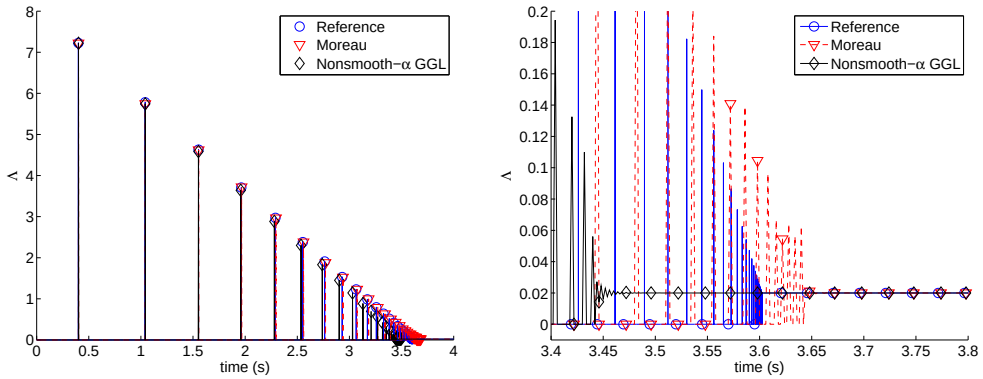

Figure 6: Bouncing ball: impulse of the contact reaction force Λ and zoom.

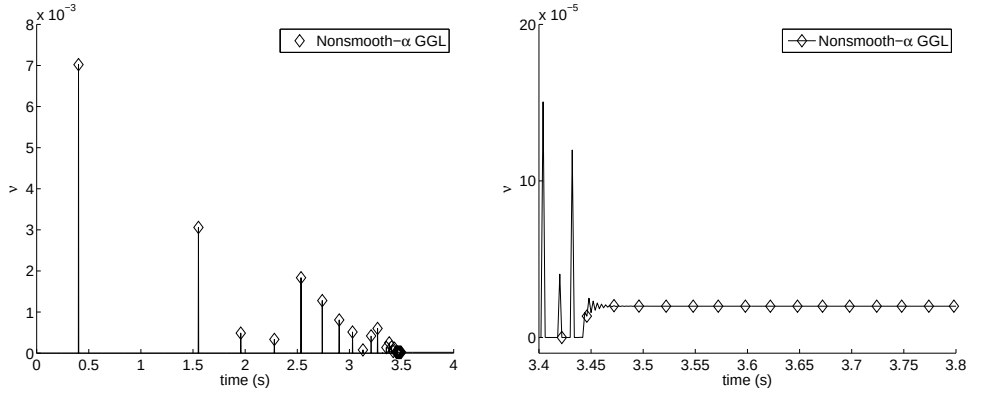

Figure 7: Bouncing ball: Lagrange multiplier ν and zoom.

As can be seen in Figure 4, the Moreau-Jean scheme and the Nonsmooth-α GGL scheme based on Eq. (28) provide a consistent evaluation of the dynamic response and of the accumulation phenomenon. However, the correct enforcement of the constraint at position level is only ensured by the Nonsmooth-α GGL scheme. Figure 5 shows that the Moreau and Nonsmooth-α GGL algorithms can represent jumps in velocities. The energy of the numerical response is a kind of staircase function, which is almost constant during the free-flight phases (it is actually strictly constant for the Nonsmooth-α GGL responses, but slowly decreasing for the Moreau response) and undergoes sudden drops at the impact times.

The impulse of the contact reaction force Λ is depicted in Figure 6. One observes some impulsive contributions at each impact instant and a continuous contribution after the accumulation when the contact is closed. The Lagrange multiplier ν, which is shown in Figure 7, can only take positive values. After the accumulation, Λ converges to ma_gh, so that Λ/h represents the reaction force, in agreement with the discussion in Section 7.1.

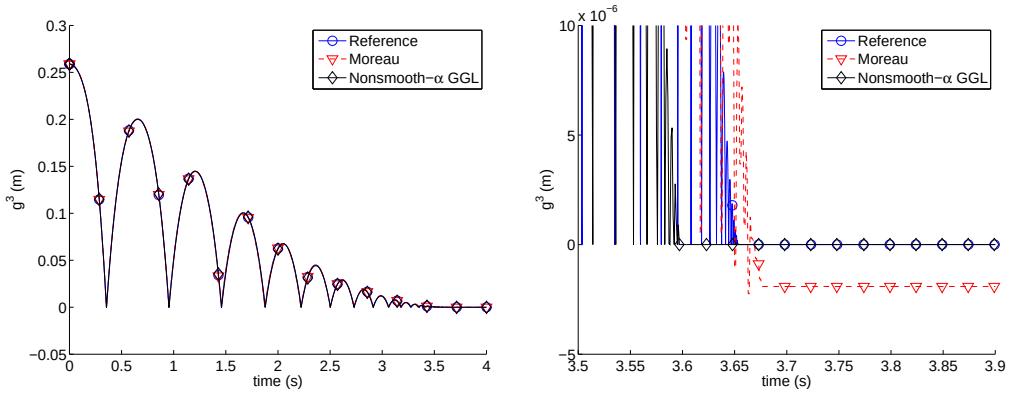

Figure 8: Bouncing pendulum: gap distance and zoom.

10.2. The bouncing pendulum

This example studies the impact of a simple rigid pendulum with both unilateral and bilateral constraints. The pendulum is constrained to swing around a pivot in the x-y plane. It consists of a massless rod and a concentrated mass at the tip. The kinematics of the simple pendulum is described in a redundant manner in terms of three Cartesian coordinates $\mathbf{q} = [x \quad y \quad \theta]^T$. Since the system has one degree-of-freedom, two nonlinear bilateral constraints are expressed as $\mathbf{g}^{\overline{\mathcal{U}}} = [x - l\cos\theta \quad y - l\sin\theta]^T = \mathbf{0}$ with $\overline{\mathcal{U}} = \{1, 2\}$. The pendulum is subjected to the gravity acceleration a_g, and thus the vector of external forces reads: $\mathbf{f}^{ext} = [0 \quad -ma_g \quad 0]^T$. The mass matrix is $\mathbf{M} = \mathrm{diag}\,[m \quad m \quad J]$.

As in [16], the physical parameters of the model are the mass $m = 1$ kg, the moment of inertia $J = 0.1$ kgm^2, the length $l = 1$ m, the gravity acceleration $a_g = 10$ m/s^2, the initial position $\mathbf{q}_0 = [l\cos\theta_0 \quad l\sin\theta_0 \quad \theta_0]^T$, with $\theta_0 = \pi/12$ rad. The pendulum is released from rest and swings clockwise due to the gravity force. An obstacle is placed such that a unilateral constraint is enforced as $0 \leq g^{\mathcal{U}} = x - \sqrt{2}/2 \perp \mathrm{d}i^{\mathcal{U}} \geq 0$ with $\mathcal{U} = \{3\}$. The coefficient of restitution is chosen as $e = 0.8$. In the numerical tests, the nominal time step size is $h = 10^{-3}$ s and the reference solution is obtained using the Nonsmooth-α GGL method and a smaller time step size $h = 10^{-4}$ s.

Since impacts occur in the dynamic response, the Smooth-α index-3 algorithm is not considered here. As in the bouncing ball example, we observe in Figures 8 and 9 that both the Moreau and the Nonsmooth-α GGL method lead to a consistent representation of jumps in velocities and of the accumulation phenomenon in the dynamic response. Also, the unilateral constraint at position level is exactly satisfied by the Nonsmooth-α GGL scheme, but not by the Moreau method. In Figure 10, the impulse of the contact reaction force includes impulsive and continuous contributions. When the contact is closed, after the accumulation phenomenon, the contact force can be estimated as Λ^3/h. This is consistent with the observation that, in that phase, Λ^3 is 10 times smaller for the reference solution, since it has been obtained using a 10 times smaller time step.

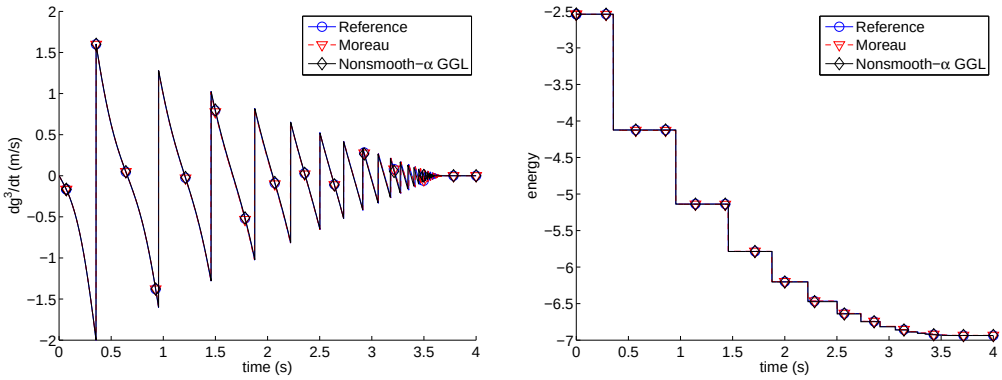

Figure 9: Bouncing pendulum: velocity and energy.

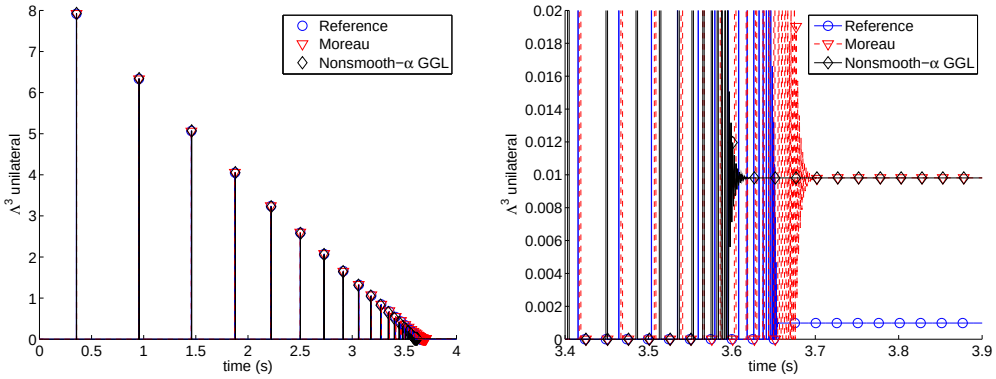

Figure 10: Bouncing pendulum: unilateral multiplier Λ^3 and zoom.

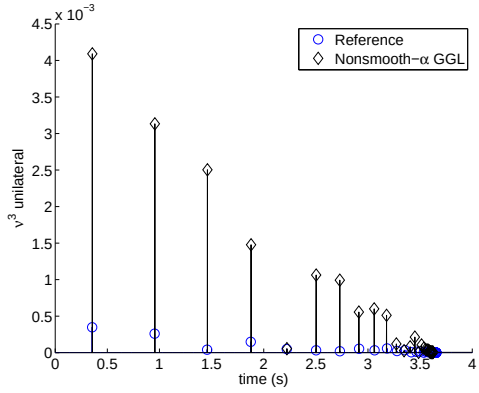

Figure 11: Bouncing pendulum: unilateral multiplier ν^3.

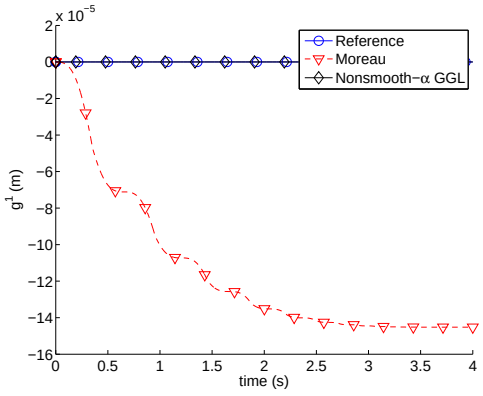

Figure 12: Bouncing pendulum: bilateral constraint at position level.

The Lagrange multiplier ν^3 is shown in Figure 11. The value of ν^3 is smaller for the reference solution, which is consistent with the fact that this multiplier is expected to converge to zero when the time step decreases.

Let us study the bilateral constraints. Figure 12 represents the first bilateral constraint at position level. A drift phenomenon is observed for the Moreau method. In contrast, the bilateral constraint is satisfied exactly at position level (as well as at velocity level) by the Nonsmooth-α GGL method. The Lagrange multipliers of the bilateral constraints are shown in Figure 13. In the Nonsmooth-α GGL scheme, Λ^1 differs from zero only if the unilateral constraint is active, whereas $\tilde{\lambda}^1$ has permanently a non-zero value. Some spurious transient oscillations are observed for $\tilde{\lambda}^1$ after each impact but they are quickly damped out. These oscillations can be explained by the fact that, between impacts, the dynamic response is computed by solving an index-2 GGL formulation of a DAE with disturbed initial values. Indeed, after the impact, the algorithm only guarantees that the constraints are satisfied at position and velocity levels, but not at acceleration-level. The violation of the hidden constraints at acceleration-level after an impact leads to transient numerical oscillations which are damped out thanks to the presence of numerical dissipation in the generalized-α algorithm, see [26] for a more detailed analysis of this algorithm for DAEs. The importance of these disturbances decreases when the time step is decreased or when the spectral radius is decreased, see Figure 14. For a spectral radius $\rho = 0$, the oscillations are completely eliminated from the numerical solution. Even though these oscillations are not desirable, in our opinion, they can be tolerated in practical applications. Let us analyze the multiplier Λ^1 obtained using the Moreau–Jean scheme (there is no smooth multiplier $\tilde{\lambda}^1$ in this case). Between the impacts and when the contact is closed, this Lagrange multiplier has a continuous component, but impulses occur at every impact time. It appears that these impulses match quite well the value of Λ^1 obtained from the Nonsmooth-α GGL scheme, and that the continuous part has some similarities with the value $h\tilde{\lambda}^1$. According to Section 7.1, when the contact is either closed or inactive, it is more appropriate to

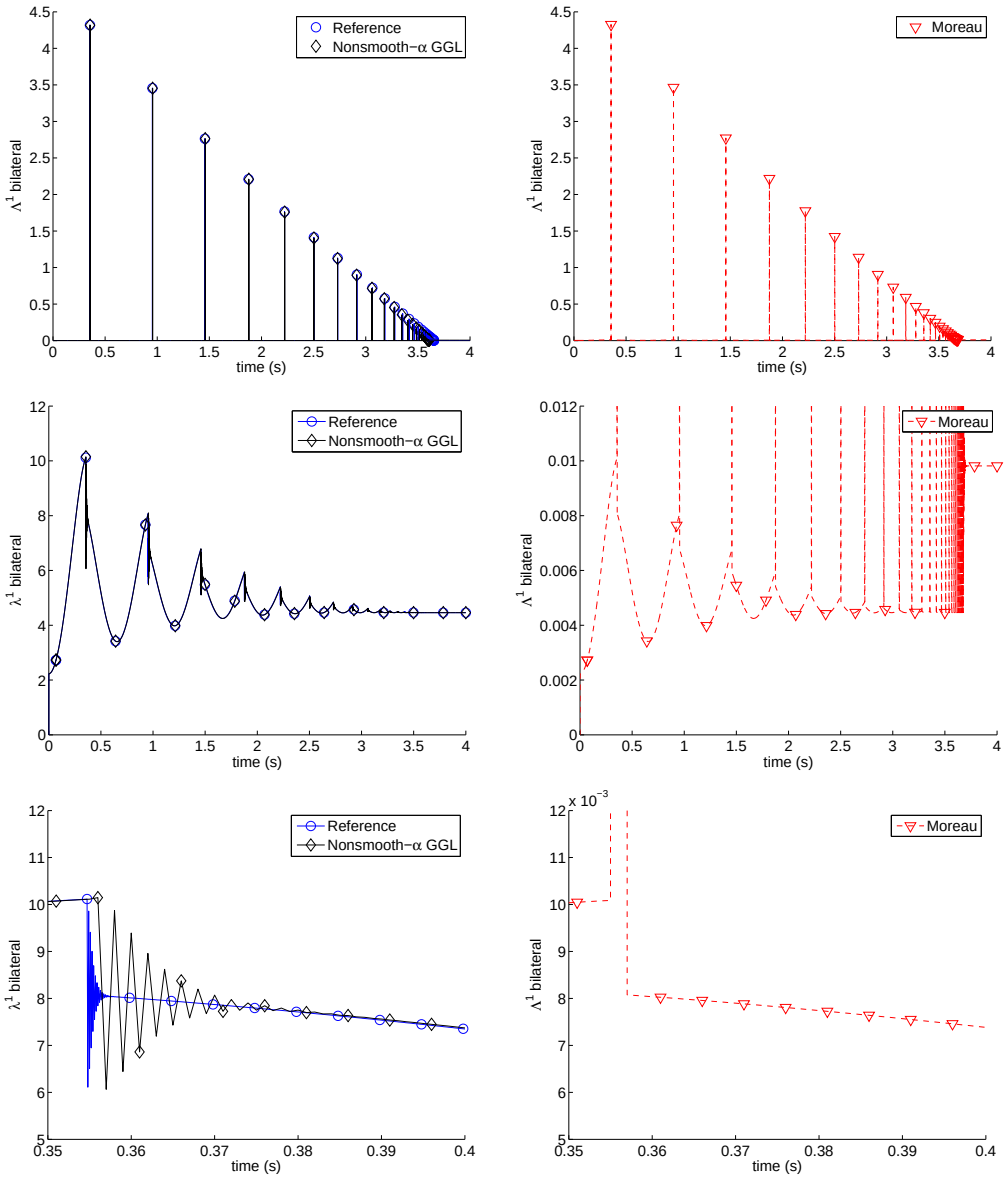

Figure 13: Bouncing pendulum: bilateral multipliers (first column: Λ^1, $\tilde{\lambda}^1$ and zoom on $\tilde{\lambda}^1$; second column: Λ^1, zoom 1 on Λ^1 and zoom 2 on Λ^1).

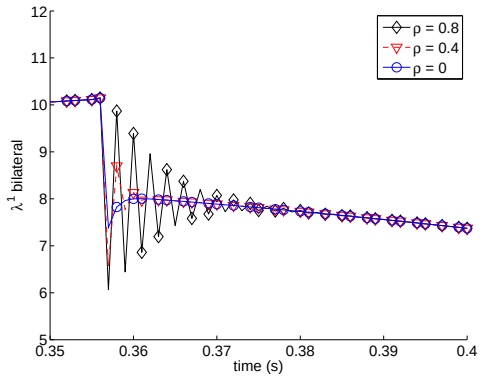

Figure 14: Bouncing pendulum: bilateral multiplier $\tilde{\lambda}^1$ for three different values of the spectral radius in the Nonsmooth-α GGL method.

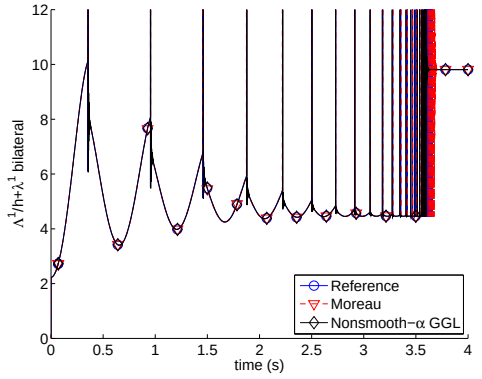

Figure 15: Bouncing pendulum: bilateral reaction force $\Lambda^1/h + \tilde{\lambda}^1$.

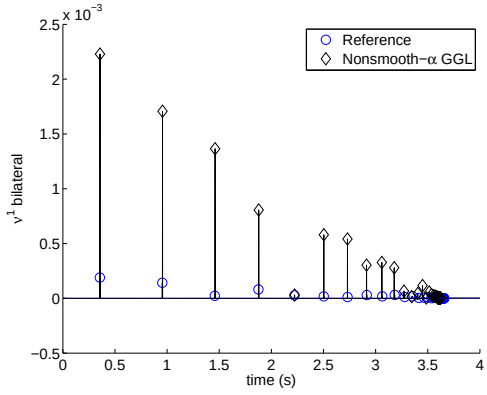

Figure 16: Bouncing pendulum: bilateral multiplier ν^1.

compare the reaction force $\Lambda^1/h + \tilde{\lambda}^1$ as shown in Figure 15, which indicates that the Moreau–Jean and the Nonsmooth-α GGL algorithms lead to similar results. Finally, the multiplier associated with the bilateral constraint at position level ν^1 is depicted in Figure 16 and similar observations can be made as in Figure 11.

10.3. Horizontal impact of an elastic bar

The impact of an elastic bar on a rigid wall is now analyzed. The bar starts its motion from a distance d_0 in an undeformed configuration with a uniform initial velocity field v_0, and bounces back against the wall. The impact is horizontal so that no gravity needs to be considered. An analytically-exact solution for this benchmark is detailed in [27]. Before the impact, the bar stays undeformed. Once the bar reaches the wall, a shock wave is caused by the impact and travels with a speed $\sqrt{E/\rho}$ to the other extremity, and then travels back to the contact point. During the travelling of the shock wave, the bar stays in closed contact with the wall due to its compression during a period of time $\Delta t = 2L\sqrt{\rho/E}$. As soon as the wave reaches the contact point, the bar takes off and the contact is released.

The same parameters as in [27] are applied in this test example: Young Modulus $E = 900$ Pa, density $\rho = 1$ kg/m^3, undeformed initial length $L = 10$ m, initial distance from the wall $d_0 = 5.005$ m, and initial velocity $v_0 = 10$ m/s. In this case, the bar stays in closed contact during a period of time $\Delta t = 2/3$ s. The numerical responses are compared with an analytical solution.

A restitution coefficient between the flexible rod and the ground needs to be defined. For flexible bodies, the restitution coefficient does not have a clear physical meaning but it should rather be considered as a numerical damping coefficient which dissipates a small amount of energy at the contact point when an impact occurs. For this reason, its value is chosen as $e = 0$.

This one-dimensional problem is discretized using 200 finite elements. The time step size is chosen as $h = 2.\,10^{-3}$ s, so that the Courant number is 1.2.

Let us analyze the numerical response. Figure 17 shows the position at the bottom of the bar. The three methods lead to reasonable results and are able to represent to closed contact during a finite time interval. However, one observes some oscillations for the Smooth-α index-3 algorithm and some penetration for the Moreau method. The velocity and the total energy of the bouncing elastic bar are analyzed in Figure 18. For the Smooth-α index-3 algorithm, the velocity constraint is not exactly satisfied when the contact is closed and the energy evolves in a non-monotonous way, which indicates a lack of consistency. In contrast, the constraint at velocity level is satisfied by the Moreau and the Nonsmooth-α GGL methods when the contact is closed. The Nonsmooth-α GGL method leads to less energy dissipation after the impact than the Moreau method, so that one can conclude that the Nonsmooth-α GGL scheme leads to a better energy representation than the Moreau–Jean scheme. The unilateral Lagrange multipliers λ (Smooth-α scheme) and Λ (Moreau–Jean and Nonsmooth-α schemes) are depicted in Figure 19 and the contact reaction force can then be derived either as λ or as Λ/h. Here, the Smooth-α method leads to significant spurious numerical

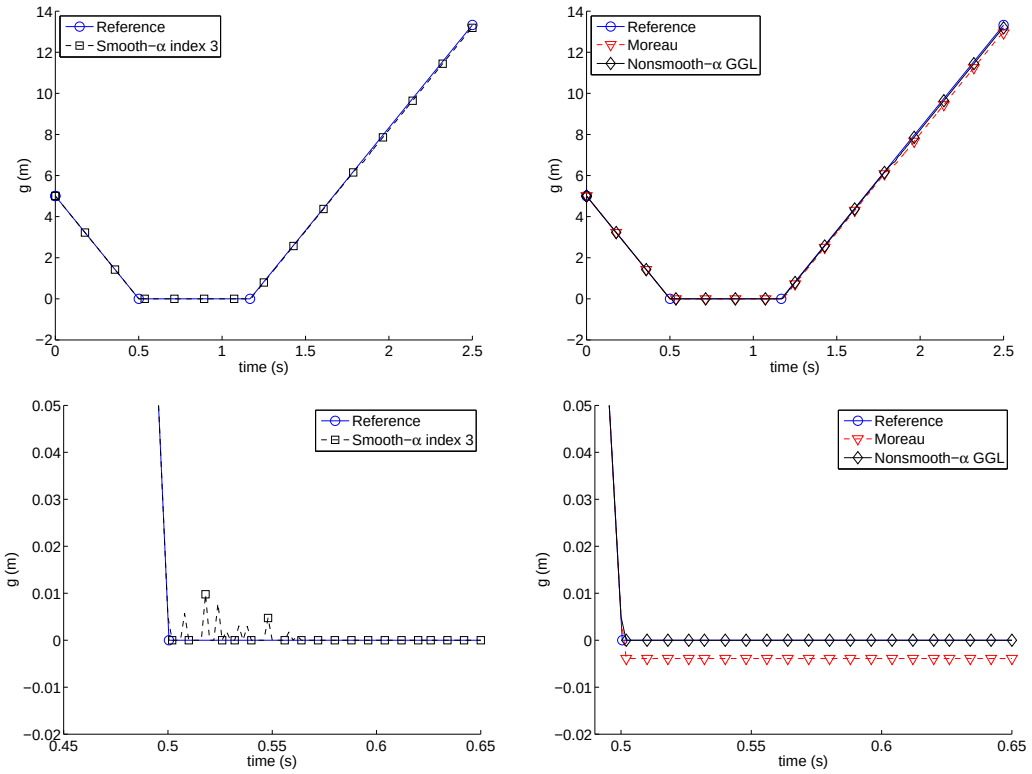

Figure 17: Bar impact: gap distance and zoom.

oscillations, while the numerical solution is much more stable for the Moreau–Jean and the Nonsmooth-α GGL method. Therefore, the Moreau–Jean and the Nonsmooth-α schemes lead to a significantly better prediction of the contact force than the Smooth-α method.

10.4. Rocking block

The rocking block example shown in Fig. 1(d) is also analyzed in [12]. The configuration is represented using three generalized coordinates $\mathbf{q} = [x\ y\ \theta]^T$, i.e., the horizontal and vertical positions of the center of mass and the orientation of the block. The equations of motion include two nonlinear unilateral constraints

$$\mathbf{M} = \begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & J \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} 0 \\ -ma_g \\ 0 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} y - \frac{l}{2}\cos\theta + \frac{L}{2}\sin\theta \\ y - \frac{l}{2}\cos\theta - \frac{L}{2}\sin\theta \end{bmatrix} \tag{53}$$

and the constraint gradient is computed as

$$\mathbf{g}_\mathbf{q} = \begin{bmatrix} 0 & 1 & \frac{l}{2}\sin\theta + \frac{L}{2}\cos\theta \\ 0 & 1 & \frac{l}{2}\sin\theta - \frac{L}{2}\cos\theta \end{bmatrix} \tag{54}$$

The system undergoes multiple impacts and an accumulation phenomenon is observed. The numerical parameters are as in [12], with $m = 1$, $L = 1$, $l = 1.5$,

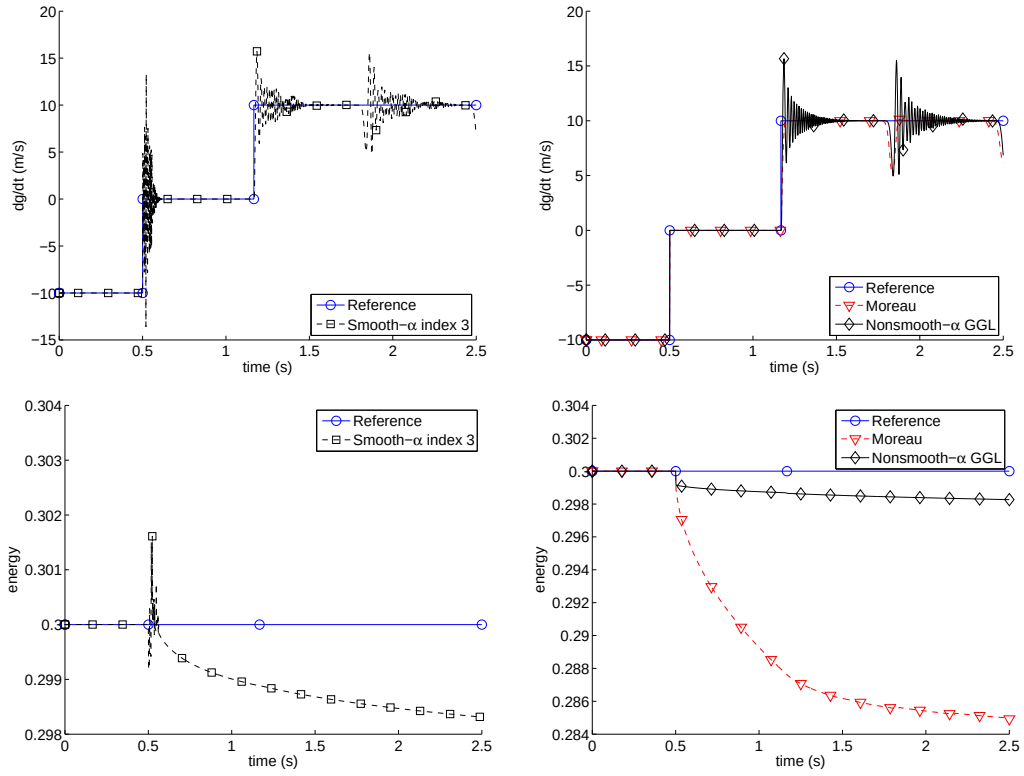

Figure 18: Bar impact: velocity and energy.

$J = m(L^2 + l^2)/12$, $a_g = 9.81$, the coefficient of restitution is $e = 0.5$ for both constraints and the initial values are $\mathbf{q}_0 = [0\ 1\ 0.2]^T$, $\mathbf{v}_0 = \mathbf{0}$. The nominal time step is $h = 10^{-2}$ s and the reference solution is obtained using the Nonsmooth-α GGL scheme and a time step $h = 2.\,10^{-4}$ s.

The Moreau scheme and the Nonsmooth-α GGL method provide a satisfactory dynamic response, which is shown in Figures 20-24. Similar observations are made as in the preceding examples. The Nonsmooth-α GGL scheme satisfies the constraints both at position and velocity level and it leads to an improved description of the energy behaviour, with less numerical dissipation than in the Moreau scheme.

11. Conclusion

This paper proposes a unified and valid formulation for frictionless contact conditions between rigid bodies as well as between flexible bodies for the dynamic analysis of multibody systems. It can naturally represent a wide spectrum of situations such as finite contact durations, multiple contacts, contact force impulses, velocity jumps and accumulation phenomena. The stability and the robustness of the scheme with respect to the choice of numerical parameters have been illustrated in a number of numerical examples.

Based on a GGL formulation, the algorithm guarantees the exact satisfaction

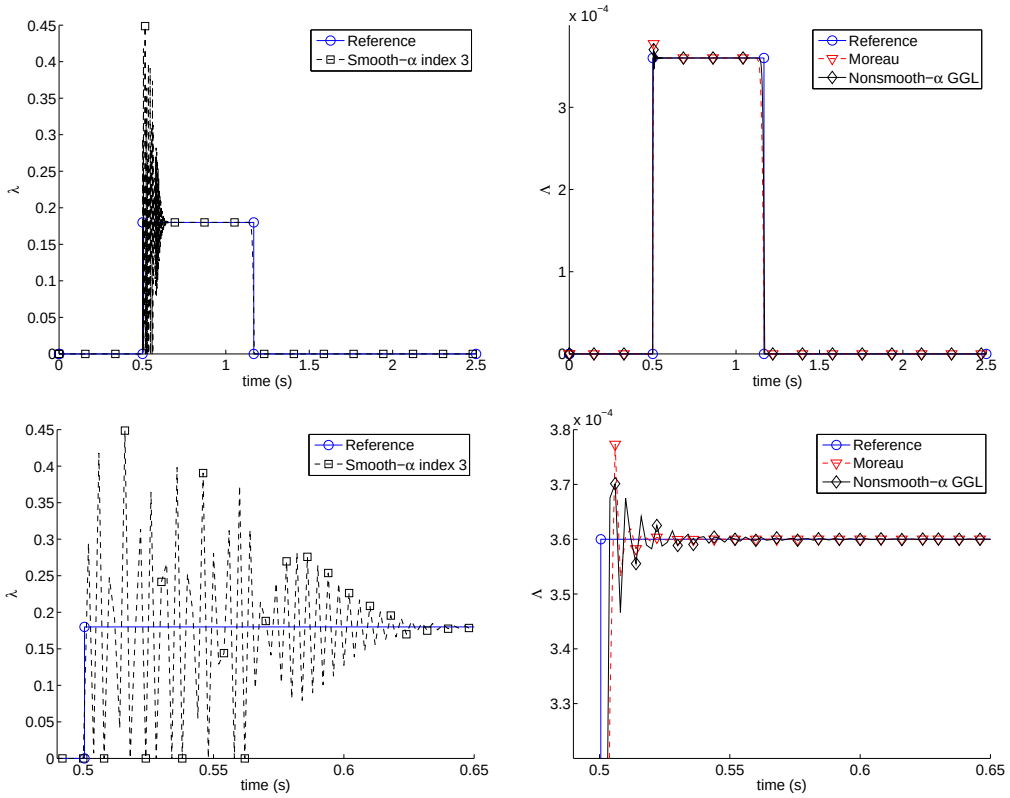

Figure 19: Bar impact: unilateral Lagrange multiplier Λ and zoom.

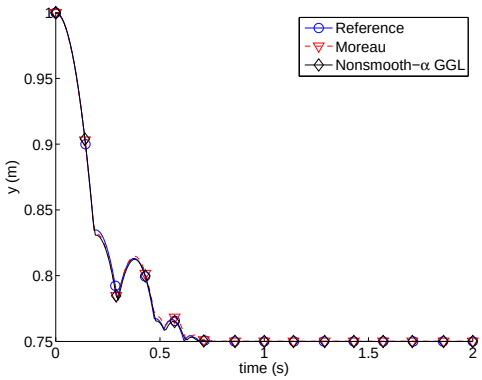

Figure 20: Rocking block: vertical position.

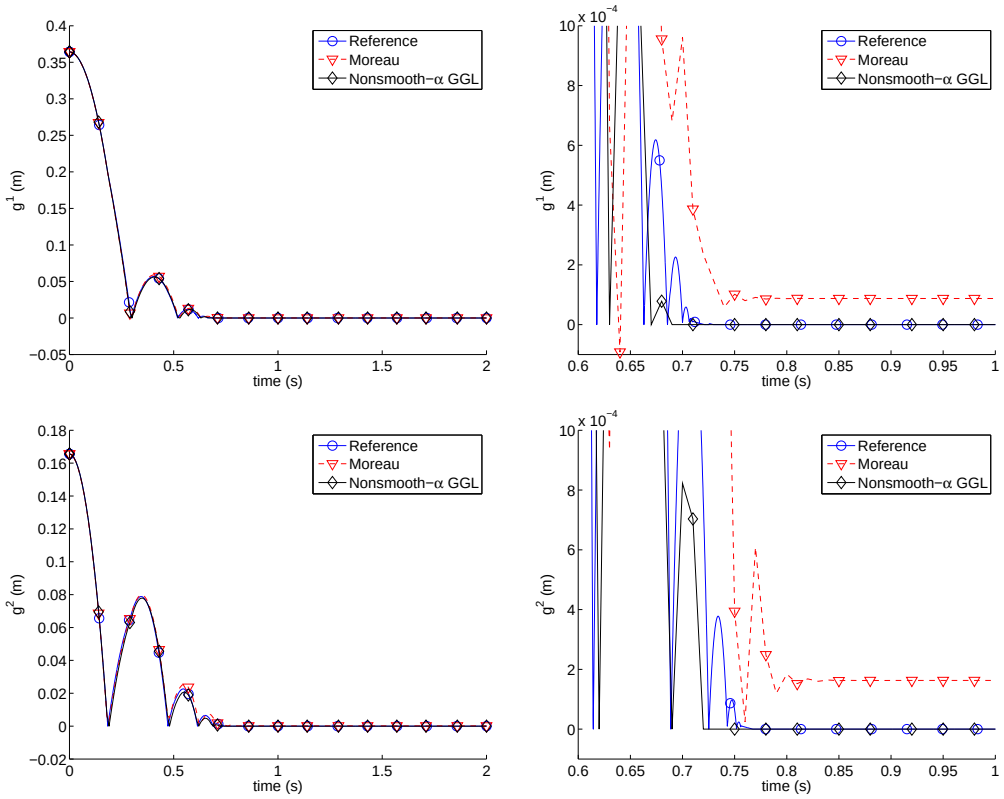

Figure 21: Rocking block: unilateral constraints and zoom.

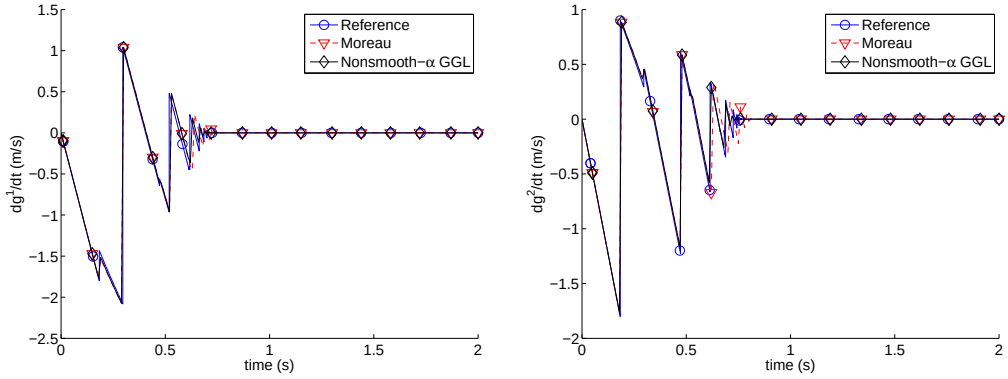

Figure 22: Rocking block: time derivative of the unilateral constraints.

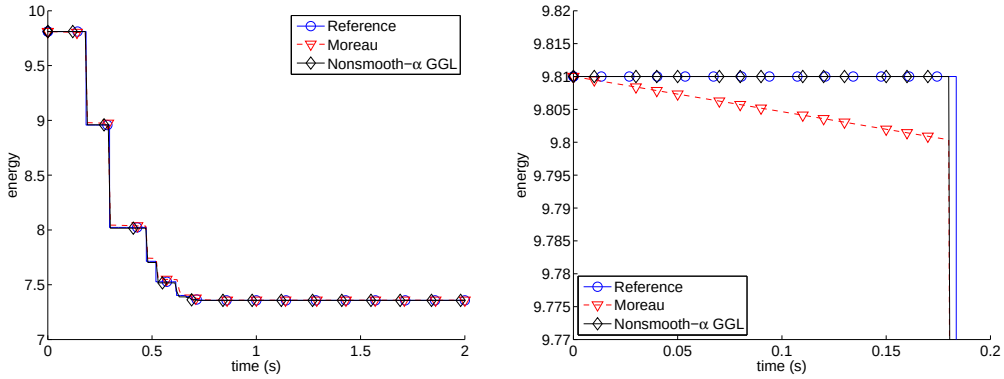

Figure 23: Rocking block: energy and zoom.

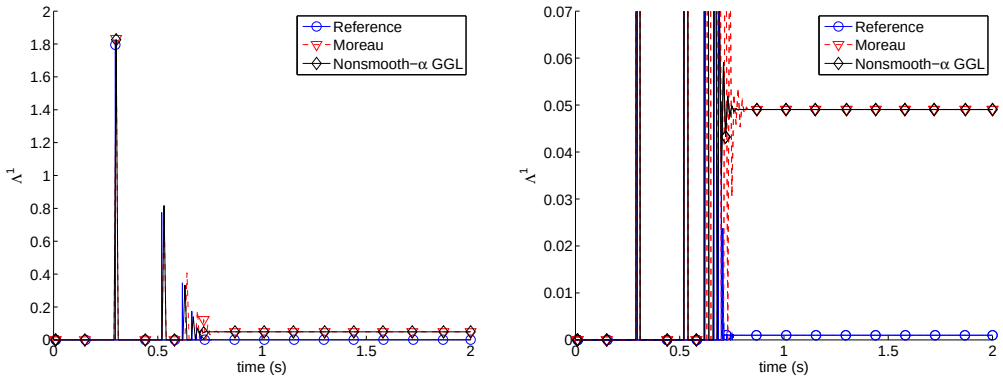

Figure 24: Rocking block: unilateral Lagrange multiplier Λ^1 and zoom.

of the complementarity condition at position level, which means that no penetration is tolerated, and at velocity level. After time discretization, the equations of motion involve two complementarity conditions and the problem can be solved efficiently at each time step using a monolithic Newton semi-smooth method. The complementarity conditions are imposed using an active set method, i.e., the activation criteria are evaluated at each Newton iteration in a fully implicit way. The numerical examples demonstrate the fast and robust convergence of the Newton semi-smooth procedure.

The algorithm also relies on a splitting of the equations to isolate smooth terms which are integrated with a higher-order accuracy than the impulsive terms. Using a generalized-α method for the smooth parts of motion, an accurate description of vibrations phenomena in nonsmooth dynamic systems can be obtained with a controllable numerical dissipation. For a given spectral radius at infinite frequencies, the numerical dissipation of the proposed Nonsmooth-α GGL scheme in the low- and mid-frequency ranges is significantly lower than the numerical dissipation of the Moreau–Jean θ scheme. Therefore, the proposed method leads to a dynamic response with a qualitatively improved energy behaviour.

As a perspective to this work, the formulation could be tested for large scale finite element models and it could be extended to frictional contact problems.

References

[1] T. Laursen, V. Chawla, Design of energy conserving algorithms for frictionless dynamic contact problems, International Journal for Numerical Methods in Engineering 40 (1997) 863–886.

[2] V. Chawla, T. Laursen, Energy consistent algorithms for frictional contact problem, International Journal for Numerical Methods in Engineering 42 (1998) 799–827.

[3] Y. Ayyad, M. Barboteu, Formulation and analysis of two energy-consistent methods for nonlinear elastodynamic frictional contact problems, Journal of Computational and Applied Mathematics 228 (2009) 254–269.

[4] D. Stewart, Rigid-body dynamics with friction and impact, SIAM Review 42 (2000) 3–39.

[5] L. Paoli, M. Schatzman, A numerical scheme for impact problems I: The one-dimensional case, SIAM Journal of Numerical Analysis 40 (2002) 702–733.

[6] L. Paoli, M. Schatzman, A numerical scheme for impact problems II: The multi-dimensional case, SIAM Journal of Numerical Analysis 40 (2002) 734–768.

[7] M. Jean, J.-J. Moreau, Dynamics in the presence of unilateral contacts and dry friction: a numerical approach, in: G. Del Pietro, F. Maceri (Eds.), Unilateral problems in structural analysis. II, CISM 304, Springer Verlag, 1987, pp. 151–196.

[8] J.-J. Moreau, Unilateral contact and dry friction in finite freedom dynamics, in: J.-J. Moreau, P. Panagiotopoulos (Eds.), Non-smooth mechanics and applications, volume 302, Springer-Verlag, Wien - New York, 1988, pp. 1–82.

[9] M. Jean, The non-smooth contact dynamics method, Computer Methods in Applied Mechanics and Engineering 177 (1999) 235–257.

[10] R. Krause, M. Walloth, A family of space-time connecting discretization schemes for elastodynamic contact problems with local impact detection, International Journal for Numerical Methods in Engineering 200 (2011) 3425–3438.

[11] C. Gear, B. Leimkuhler, G. Gupta, Automatic integration of Euler-Lagrange equations with constraints, Journal of Computational and Applied Mathematics 12–13 (1985) 77–90.

[12] V. Acary, Projected event-capturing time-stepping schemes for nonsmooth mechanical systems with unilateral contact and Coulomb's friction, Computer Methods in Applied Mechanics and Engineering 256 (2013) 224–250.

[13] S. Schoeder, H. Ulbrich, T. Schindler, Discussion on the Gear-Gupta-Leimkuhler method for impacting mechanical systems, Multibody Systems Dynamics 31 (2013) 477–495.

[14] M. Hintermüller, K. Ito, K. Kunish, The primal-dual active set strategy as a semismooth Newton method, SIAM Journal on Optimization 13 (2003) 865–888.

[15] I. Ben Gharbia, J. Gilbert, Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix, Mathematical Programming, Series A 134 (2012) 349–364.

[16] Q.-z. Chen, V. Acary, G. Virlez, O. Brüls, A nonsmooth generalized-α scheme for flexible multibody systems with unilateral constraints, International Journal for Numerical Methods in Engineering 96 (2013) 487–511.

[17] J.-J. Moreau, Numerical aspects of the sweeping process, Computer Methods in Applied Mechanics and Engineering 177 (1999) 329–349.

[18] K. Brenan, S. Campbell, L. Petzold, Numerical Solution of Initial–Value Problems in Differential–Algebraic Equations, 2nd ed., SIAM, Philadelphia, 1996.

[19] V. Acary, B. Brogliato, Numerical methods for nonsmooth dynamical systems - Applications in Mechanics and Electronics, volume 35 of *Lecture Notes in Applied and Computational Mechanics*, Springer-Verlag, Berlin, 2008.

[20] V. Acary, Higher order event capturing time–stepping schemes for nonsmooth multibody systems with unilateral constraints and impacts, Applied Numerical Mathematics 62 (2012) 1259– 1275.

[21] M. Arnold, O. Brüls, Convergence of the generalized-α scheme for constrained mechanical systems, Multibody System Dynamics 18 (2007) 185–202.

[22] J. Chung, G. Hulbert, A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-α method, ASME Journal of Applied Mechanics 60 (1993) 371–375.

[23] S. Erlicher, L. Bonaventura, O. Bursi, The analysis of the generalized-α method for non-linear dynamic problems, Computational Mechanics 28 (2002) 83–104.

[24] P. Alart, A. Curnier, A mixed formulation for frictional contact problems prone to Newton like solution methods, Computer Methods in Applied Mechanics and Engineering 92 (1991) 353–375.

[25] C. Bottasso, O. Bauchau, A. Cardona, Time-step-size-independent conditioning and sensitivity to perturbations in the numerical solution of index three differential algebraic equations, SIAM Journal on Scientific Computing 29 (2007) 397–414.

[26] M. Arnold, O. Brüls, A. Cardona, Error analysis of generalized-α Lie group time integration methods for constrained mechanical systems, Numerische Mathematik in press (2014).

[27] D. Doyen, A. Ern, S. Piperno, Time-integration schemes for the finite element dynamic Signorini problem, SIAM Journal on Scientific Computing 33 (2011) 223–249.